

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)



14 MARS 1997

BREVET D'INVENTION

CERTIFICAT D'UTILITÉ - CERTIFICAT D'ADDITION

PRIORITY DOCUMENT

REC'D 24 MAR 1997
WIPO PCT

COPIE OFFICIELLE

Le Directeur général de l'Institut national de la propriété industrielle certifie que le document ci-annexé est la copie certifiée conforme d'une demande de titre de propriété industrielle déposée à l'Institut.

Fait à Paris, le 10 FEV. 1997

Pour le Directeur général de l'Institut
national de la propriété industrielle
Le Chef de Division

Yves CAMPENON

INSTITUT
NATIONAL DE
LA PROPRIÉTÉ
INDUSTRIELLE

SIEGE
26 bis, rue de Saint Petersburg
75800 PARIS Cédex 08
Telephone : 01 53 04 53 04
Télécopie : 01 42 93 59 30

REQUETE
EN DÉLIVRANCE D'UN
TITRE DE PROPRIÉTÉ
INDUSTRIELLE *

DATE DE REMISE DES PIÈCES
02 FEV 1996

N° D'ENREGISTREMENT NATIONAL
96 01309 -

CODE POSTAL DU LIEU DE DÉPÔT
75

a ☒ BREVET D'INVENTION
b ☐ CERTIFICAT D'UTILITÉ
c ☐ DEMANDE DIVISIONNAIRE
d ☐ TRANSFORMATION D'UNE DEMANDE DE BREVET EUROPÉEN

Pour c et d, précisez : Nature, N° et date de la demande initiale

DATE DE DÉPÔT
02 FEV. 1996

4 NUMÉRO DU POUVOIR PERMANENT

2 OPTIONS OBLIGATOIRES au moment du dépôt (sauf pour le certificat d'utilité)

LE DEMANDEUR REQUIERT L'ÉTABLISSEMENT DIFFERÉ DU RAPPORT DE RECHERCHE

☐ OUI
☒ NON

SI L'OPTION CHOISIE EST NON ET SI LE DEMANDEUR EST UNE PERSONNE PHYSIQUE IL REQUIERT LE PAIEMENT ÉCHELONNÉ DE LA REDEVANCE DE RAPPORT DE RECHERCHE

☐ OUI
☐ NON

NATURE

NUMÉRO

DATE DE LA DEMANDE INITIALE

3 NOM ET ADRESSE DU DEMANDEUR OU DU MANDATAIRE à qui toute la correspondance doit être adressée

CABINET LAVOIX
2 Place d'Estienne d'Orves
75441 PARIS CEDEX 09

5 RÉFÉRENCE DU CORRESPONDANT
BFF 95/328

6 TÉLÉPHONE DU CORRESPONDANT
48 74 92 22

7 TITRE DE L'INVENTION

Nouvelles séquences d'acides nucléiques de la famille des facteurs suppresseurs de tumeurs, produits d'expression correspondants et leurs applications.

8 DEMANDEUR(S) : Nom et Prénoms (souligner le nom patronymique) ou dénomination et forme juridique

N° SIREN

Société dite : **SANOFI**

9 ADRESSE(S) COMPLÈTE(S)

32-34, rue Marbeuf 75008 PARIS

PAYS

FR

10 NATIONALITÉ(S)
Française

☒ DE DÉPÔT

REDEVANCES VERSÉES

☒ DE RAPPORT DE RECHERCHE

☐ DE REVENDICATION DE PRIORITÉ

☒ DE REVENDICATION (à partir de la 116)

11 INVENTEUR(S)

LE DEMANDEUR EST L'UNIQUE INVENTEUR

☐ OUI
☒ NON

Si la réponse est non voir notice explicative

12

SI LE DEMANDEUR EST UNE PERSONNE PHYSIQUE NON IMPOSABLE IL REQUIERT OU A REQUIS LA RÉDUCTION DES REDEVANCES

☐ OUI
☐ NON

13 DÉCLARATION DE PRIORITÉ
OU REQUÊTE DU BÉNÉFICE DE
LA DATE DE DÉPÔT D'UNE
DEMANDE ANTÉRIEURE

PAYS D'ORIGINE

DATE DE DÉPÔT

NUMÉRO

14

DIVISIONS

ANTÉRIEURES À LA PRÉSENTE DEMANDE

N°

N°

N°

N°

15 SIGNATURE DU DEMANDEUR OU DU MANDATAIRE
NOUVELLE QUANTITÉ DE SIGNATURES À L'INSCRIPTION

CABINET LAVOIX
M. BOLENSKY n° 92.1186

SIGNATURE DU PRÉPOSÉ À LA RÉCEPTION

SIGNATURE APRÈS ENREGISTREMENT DE LA DEMANDE À L'INPI



BREVET D'INVENTION, CERTIFICAT D'UTILITE

DÉSIGNATION DE L'INVENTEUR

(si le demandeur n'est pas l'inventeur ou l'unique inventeur)

DIVISION ADMINISTRATIVE DES BREVETS

26bis, rue de Saint-Petersbourg
75800 Paris Cédex 08
Tél. : (1) 42 94 52 52 - Télécopie : (1) 42 93 59 30

N° D'ENREGISTREMENT NATIONAL

96 01 309

TITRE DE L'INVENTION : Protéine purifiée SR-p70.

LE (S) SOUSSIGNÉ (S)

SANOFI

32-34, rue Marbeuf 75008 PARIS FRANCE

DÉSIGNE (NT) EN TANT QU'INVENTEUR (S) (indiquer nom, prénoms, adresse et souligner le nom patronymique) :

CAPUT Daniel
La Bousquièrre
31290 AVIGNONET LAURAGAIS FRANCE

FERRARA Pascual
Libouille Saint-Assiscle
31290 AVIGNONET LAURAGAIS FRANCE

KAGHAD Ahmèd Mourad
5 rue de la Poste
31450 MONTGISCARD FRANCE

NOTA : A titre exceptionnel, le nom de l'inventeur peut être suivi de celui de la société à laquelle il appartient (société d'appartenance) lorsque celle-ci est différente de la société déposante ou titulaire.

Date et signature (s) du (des) demandeur (s) ou du mandataire

Paris, le 29 Janvier 1997

CABINET LAVOIX
M. OBOLENSKY n° 92.1185

DOCUMENT COMPORTANT DES MODIFICATIONS

PAGE(S) DE LA DESCRIPTION OU DES REVENDICATIONS OU PLANCHE(S) DE DESSIN			R.M.*	DATE DE LA CORRESPONDANCE	TAMPON DATEUR DU CORRECTEUR
Modifiée(s)	Supprimée(s)	Ajoutée(s)			
54, 55			X	06/06/96	1. JUIN 1996 . 01
54		56 57	X	26/12/96	

Un changement apporté à la rédaction des revendications d'origine, sauf si celui-ci découle des dispositions de l'article 28 du décret du 19 septembre 1979, est signalé par la mention "R.M." (revendications modifiées).

L'invention concerne de nouvelles séquences d'acides nucléiques de la famille des gènes suppresseurs de tumeurs apparentée avec le gène de la protéine p53, et les séquences protéiques correspondantes.

5 L'invention concerne également les applications prophylactiques, thérapeutiques et diagnostiques de celles-ci, notamment dans le domaine des pathologies liées aux phénomènes d'apoptose ou de transformation cellulaire.

10 Les gènes suppresseurs de tumeurs jouent un rôle clef dans la protection contre les phénomènes de cancérisation, et toute modification susceptible d'entraîner la perte de l'un de ces gènes, son inactivation ou son dysfonctionnement, peut avoir un caractère oncogène, créant ainsi des conditions favorables au développement d'un cancer.

15 Les auteurs de la présente invention ont identifié les produits de transcription d'un nouveau gène ainsi que les protéines correspondantes. Ce gène SR-p70, est apparenté au gène suppresseur de tumeur p53, dont l'activité anti-tumorale est liée à son activité de facteur de transcription et plus spécifiquement aux contrôles exercés sur l'activité des gènes Bax et Bcl-2, instrumentaux dans les mécanismes de mort cellulaire.

20 La présente invention est donc relative à des protéines purifiées SR-p70, ou des fragments biologiquement actifs de celles-ci.

L'invention concerne également des séquences d'acides nucléiques isolées codant pour lesdites protéines ou leurs fragments biologiquement actifs et les sondes nucléotidiques obtenues à partir de ces séquences.

25 Elle vise en outre les vecteurs de clonage et/ou d'expression contenant au moins l'une des séquences nucléotidiques définies ci-dessus, et les cellules hôtes transfectées par ces vecteurs de clonage et/ou d'expression dans des conditions permettant la répllication et/ou l'expression de l'une desdites séquences nucléotidiques.

30 Les méthodes de production de protéines recombinantes SR-p70 ou de leurs fragments biologiquement actifs par les cellules hôtes transfectées font également partie de l'invention.

L'invention comprend également des anticorps ou des dérivés d'anticorps spécifiques des protéines définies ci-dessus.

35 Elle vise en outre des méthodes de détection des cancers, soit par la mesure de l'accumulation des protéines SR-p70 dans les tumeurs selon des techniques d'immuno-histochimie, soit par la mise en évidence dans le sérum de patients d'auto-anticorps dirigés contre ces protéines.

L'invention concerne également tout inhibiteur ou activateur de l'activité du SR-p70 par exemple d'interaction protéine-protéine faisant intervenir le SR-p70.

5 Elle concerne aussi des séquences oligonucléotidiques antisens pouvant moduler *in vivo* l'expression du gène SR-p70.

L'invention comprend enfin une méthode de thérapie génique dans laquelle des vecteurs tels que par exemple des vecteurs viraux inactivés capables de transférer des séquences codantes pour une protéine selon l'invention sont injectés à des cellules déficientes pour cette protéine, à des fins de régulation
10 des phénomènes d'apoptose ou de réversion de la transformation.

La présente invention a pour objet un polypeptide purifié comprenant une séquence d'acides aminés choisie parmi :

- a) la séquence SEQ ID n° 2 ;
- b) la séquence SEQ ID n° 4 ;
- 15 c) la séquence SEQ ID n° 6 ;
- d) la séquence SEQ ID n° 8 ;
- e) la séquence SEQ ID n° 10 ;
- f) toute séquence biologiquement active dérivée de SEQ ID n° 2, SEQ ID n° 4, SEQ ID n° 6, SEQ ID n° 8 ou SEQ ID n° 10.

20 Dans la description de l'invention, on utilise les définitions suivantes :

- protéine SR-p70 : un polypeptide comprenant une séquence d'acides aminés choisie parmi les séquences SEQ ID n° 2, SEQ ID n° 4, SEQ ID n° 6, SEQ ID n° 8 ou SEQ ID n° 10, ou tout fragment ou dérivé de celui-ci biologiquement actif.
- 25 - dérivé : tout polypeptide variant du polypeptide de séquence SEQ ID n° 2, SEQ ID n° 4, SEQ ID n° 6, SEQ ID n° 8 ou SEQ ID n° 10, ou toute molécule résultant d'une modification de nature génétique et/ou chimique de la séquence SEQ ID n° 2, SEQ ID n° 4, SEQ ID n° 6, SEQ ID n° 8 ou SEQ ID n° 10, c'est-à-dire obtenue par mutation, délétion, addition, substitution et/ou modification chimique d'un seul ou d'un nombre limité d'acides aminés, ainsi que toute
30 séquence isoforme, c'est-à-dire une séquence identique à la séquence SEQ ID n° 2 ou SEQ ID n° 4, SEQ ID n° 6, SEQ ID n° 8 ou SEQ ID n° 10, à l'un de ses fragments ou séquences modifiées, contenant un ou plusieurs acides aminés sous la forme d'énantiomère D, lesdites séquences variantes, modifiées ou isoformes ayant conservé au moins l'une des propriétés les rendant
35 biologiquement actives.

- , biologiquement actif : capable de se lier à l'ADN et/ou d'exercer une activité de facteur de transcription et/ou de participer au contrôle du cycle cellulaire, de la différenciation et de l'apoptose et/ou capable d'être reconnu par les anticorps spécifiques du polypeptide de séquence SEQ ID n°2, SEQ ID n°4, SEQ ID n° 6, SEQ ID n° 8 ou SEQ ID n° 10, et/ou capable d'induire des anticorps qui reconnaissent ce polypeptide.

La fabrication de dérivés peut avoir différents objectifs, dont en particulier celui d'augmenter l'affinité du polypeptide pour l'ADN ou son activité de facteur de transcription, celui d'améliorer ses taux de production, d'augmenter sa résistance à des protéases, de modifier ses activités biologiques ou de lui conférer de nouvelles propriétés pharmaceutiques et/ou biologiques.

Parmi les polypeptides de l'invention, on préfère le polypeptide d'origine humaine, comprenant la séquence SEQ ID n° 6. Ce polypeptide de 636 acides aminés est identique à plus de 97 % au polypeptide de séquence SEQ ID n° 2.

Le polypeptide de séquence SEQ ID n° 2 et celui de séquence SEQ ID n° 4 sont deux produits d'expression d'un même gène, de même pour les SEQ ID n° 8 et SEQ ID n° 10.

Comme il sera expliqué dans les exemples, le polypeptide de séquence SEQ ID n° 4 correspond à une terminaison prématurée du peptide de séquence SEQ ID n° 2, liée à un épissage alternatif du transcript codant pour le polypeptide de SEQ ID n° 2 le plus long (ARN messenger) du gène correspondant. La séquence peptidique N-terminale de la séquence SEQ ID n° 10 est déletée, ce lié à un épissage alternatif de son transcript codant, comme pour la séquence SEQ ID n° 4.-

Avantageusement, l'invention vise un polypeptide correspondant au domaine de fixation sur l'ADN de l'un des polypeptides précédents.

Ce domaine correspond à la séquence comprise entre le résidu 110 et le résidu 310 pour les séquences SEQ ID n° 2, 4 ou 6, le résidu 60 et le résidu 260 pour la séquence SEQ ID n° 8 et le résidu 109 et le résidu 123 pour la séquence SEQ ID n° 10.

La présente invention a également pour objet des séquences d'acides nucléiques codant pour une protéine SR-p70 ou des fragments ou dérivés de celle-ci biologiquement actifs.

Plus préférentiellement, l'invention a pour objet une séquence nucléotidique d'acides nucléiques isolée choisie parmi :

a) la séquence SEQ ID n° 1 ;

b) la séquence SEQ ID n° 3 ;

c) la séquence SEQ ID n° 5 ;

d) la séquence SEQ ID n° 7 ;

e) la séquence SEQ ID n° 9 ;

5 f) les séquences d'acides nucléiques capables de s'hybrider à la séquence SEQ ID n° 1, SEQ ID n° 3, SEQ ID n° 5, SEQ ID n° 7 ou SEQ ID n° 9 ou à leurs séquences complémentaires, ou de s'hybrider à leurs séquences proximales.

g) les séquences dérivées des séquences a), b), c), d), e) ou f) du fait de la dégénérescence du code génétique.

10 Selon un mode de réalisation préféré, l'invention a pour objet la séquence nucléotidique SEQ ID n° 5 correspondant à l'ADNc de la protéine humaine de séquence SEQ ID n° 6.

15 Les différentes séquences nucléotidiques de l'invention peuvent être d'origine artificielle ou non. Il peut s'agir de séquences d'ADN ou d'ARN, obtenues par criblage de banques de séquences au moyen de sondes élaborées sur la base des séquences SEQ ID n° 1, 3, 5, 7 ou 9. De telles banques peuvent être préparées par des techniques classiques de biologie moléculaire, connues de l'homme de l'art.

20 Les séquences nucléotidiques selon l'invention peuvent également être préparées par synthèse chimique, ou encore par des méthodes mixtes incluant la modification chimique ou enzymatique de séquences obtenues par criblage de banques.

25 Ces séquences nucléotidiques permettent la réalisation de sondes nucléotidiques, capables de s'hybrider spécifiquement avec une séquence d'acides nucléiques, y compris un ARN messager, codant pour un polypeptide selon l'invention ou un fragment biologiquement actif de celui-ci. De telles sondes font également partie de l'invention. Elles peuvent être utilisées comme outil de diagnostic *in vitro* pour la détection, par des expériences d'hybridation, de transcrits spécifiques des polypeptides de l'invention dans des échantillons
30 biologiques ou pour la mise en évidence de synthèses aberrantes ou d'anomalies génétiques telles que la perte d'hétérozygotie ou le réarrangement génétique, résultant d'un polymorphisme, de mutations ou d'un épissage différent.

35 Les sondes de l'invention comportent au minimum 10 nucléotides, et au maximum comportent la totalité de la séquence du gène SR-p70 contenu par exemple dans un cosmide.

Parmi les sondes les plus courtes, c'est-à-dire d'environ 10 à 20 nucléotides, les conditions d'hybridation appropriées correspondent aux conditions de température et de force ionique usuellement utilisées par l'homme du métier.

5 Avantageusement, ces sondes sont représentées par les oligonucléotides suivants ou leurs complémentaires :

SEQ ID n° 11 : GCG AGC TGC CCT CGG AG
 SEQ ID n° 12 : GGT TCT GCA GGT GAC TCA G
 SEQ ID n° 13 : GCC ATG CCT GTC TAC AAG
 10 SEQ ID n° 14 : ACC AGC TGG TTG ACG GAG
 SEQ ID n° 15 : GTC AAC CAG CTG GTG GGC CAG
 SEQ ID n° 16 : GTG GAT CTC GGC CTC C

Préférentiellement, les sondes de l'invention sont marquées, préalablement à leur utilisation. Pour cela, plusieurs techniques sont à la portée
 15 de l'homme du métier (marquage fluorescent, radioactif, chimioluminescent, enzymatique, etc).

Les méthodes de diagnostic *in vitro* dans lesquelles ces sondes nucléotidiques sont mises en oeuvre, sont incluses dans l'objet de la présente invention.

20 Ces méthodes concernent par exemple la détection de synthèses anormales (ex. accumulation de produits de transcription) ou d'anomalies génétiques, telles que la perte d'hétérozygotie et le réarrangement génétique, et les mutations ponctuelles au niveau des séquences nucléotidiques d'acides nucléiques codant pour une protéine SR-p70, selon la définition donnée
 25 précédemment.

Les séquences nucléotidiques de l'invention sont également utiles pour la fabrication et l'utilisation d'amorces oligonucléotidiques pour des réactions de séquençage ou d'amplification spécifique selon la technique dite de PCR ou toute variante de celle-ci (Ligase Chain Reaction (LCR), ...).

30 Des paires d'amorces préférées sont constituées par des amorces choisies sur les séquences nucléotidiques : SEQ ID n° 1 : séquence de singe de 2 874 nucléotides et SEQ ID n°5 : ADNc SR-p70 humain, notamment en amont du codon ATG d'initiation et en aval du codon TGA d'arrêt de traduction.

Avantageusement, ces amorces sont représentées par les couples
 35 suivants:

- couple n°1 : : SEQ ID n° 11
 amorce sens : GCG AGC TGC CCT CGG AG

5 : SEQ ID n° 12
 amorce antisens : GGT TCT GCA GGT GAC TCA G

- couple n°2 : : SEQ ID n°13
 amorce sens : GCC ATG CCT GTC TAC AAG

10 : SEQ ID n° 14
 amorce antisens : ACC AGC TGG TTG ACG GAG

- couple n° 3 : : SEQ ID n° 15
 amorce sens : GTC AAC CAG CTG GTG GGC CAG

15 : SEQ ID n° 16
 amorce antisens : GTG GAT CTC GGC CTC C

Ces amorces correspondent aux séquences allant respectivement :

20

- du nucléotide n° 124 au nucléotide n° 140 sur SEQ ID n° 1 et du nucléotide n° 1 au nucléotide n° 17 sur SEQ ID n°5 pour SEQ ID N° 11

- du nucléotide n° 2280 au nucléotide n° 2262 sur SEQ ID n° 1 et du nucléotide n° 2156 au nucléotide 2138 sur SEQ ID n°5 pour SEQ ID N° 12

25 - du nucléotide n° 684 au nucléotide n° 701 sur SEQ ID n° 1 pour SEQ ID N° 13

- du nucléotide n° 1447 au nucléotide n° 1430 sur SEQ ID n° 1 et du nucléotide 1324 au nucléotide 1307 sur SEQ ID n°5 pour SEQ ID N° 14

30 - du nucléotide 1434 au nucléotide 1454 sur SEQ ID n°1 et du nucléotide 1311 au nucléotide 1331 sur SEQ ID n°5 pour SEQ ID n° 15

- du nucléotide 2066 au nucléotide 2051 sur SEQ ID n°1 et du nucléotide 1940 au nucléotide 1925 dans SEQ ID n°5 pour SEQ ID n°16.

35

Les séquences nucléotidiques selon l'invention pourraient avoir par ailleurs des utilisations en thérapie génique, notamment pour le contrôle des phénomènes d'apoptose et de réversion de la transformation.

Les séquences nucléotidiques selon l'invention peuvent par ailleurs être utilisées pour la production de protéines recombinantes SR-p70, selon la définition qui a été donnée à ce terme.

5 Ces protéines peuvent être produites à partir des séquences nucléotidiques définies ci-dessus, selon des techniques de production de produits recombinants connues de l'homme du métier. Dans ce cas, la séquence nucléotidique utilisée est placée sous le contrôle de signaux permettant son expression dans un hôte cellulaire.

10 Un système efficace de production d'une protéine recombinante nécessite de disposer d'un vecteur, par exemple d'origine plasmidique ou virale, et d'une cellule hôte compatible.

15 L'hôte cellulaire peut être choisi parmi des systèmes procaryotes, comme les bactéries, ou eucaryotes, comme par exemple les levures, cellules d'insectes, CHO (cellules d'ovaires de hamster chinois) ou tout autre système avantageusement disponible. Un hôte cellulaire préféré pour l'expression des protéines de l'invention est constitué par la bactérie *E. coli*, notamment la souche MC 1061 (Clontec).

20 Le vecteur doit comporter un promoteur, des signaux d'initiation et de terminaison de la traduction, ainsi que les régions appropriées de régulation de la transcription. Il doit pouvoir être maintenu de façon stable dans la cellule et peut éventuellement posséder des signaux particuliers spécifiant la sécrétion de la protéine traduite.

25 Ces différents signaux de contrôle sont choisis en fonction de l'hôte cellulaire utilisé. A cet effet, les séquences nucléotidiques selon l'invention peuvent être insérées dans des vecteurs à réplication autonome au sein de l'hôte choisi, ou des vecteurs intégratifs de l'hôte choisi. De tels vecteurs seront préparés selon les méthodes couramment utilisées par l'homme du métier, et les clones en résultant peuvent être introduits dans un hôte approprié par des méthodes standard, telles que par exemple l'électroporation.

30 Les vecteurs de clonage et/ou d'expression contenant au moins l'une des séquences nucléotidiques définies ci-dessus font également partie de la présente invention.

35 Un vecteur de clonage et d'expression préféré est le plasmide PSE1 qui comporte à la fois les éléments nécessaires pour son utilisation comme vecteur de clonage dans *E. coli* (origine de réplication dans *E. coli* et gène de résistance à l'ampicilline, provenant du plasmide pTZ 18R), et comme vecteur

d'expression dans les cellules animales (promoteur, intron, site de polyadenylation, origine de réplication du virus SV40), ainsi que les éléments permettant sa copie en simple brin dans un but de séquençage (origine de réplication du phage f1).

5 Les caractéristiques de ce plasmide sont décrites dans la demande EP 0 506 574.

Sa construction, ainsi que l'intégration des ADNc provenant des séquences d'acides nucléiques de l'invention sont par ailleurs décrites dans les exemples ci-après.

10 Selon un mode de réalisation préféré, les protéines de l'invention sont sous forme de protéines de fusion, notamment sous forme de protéine fusionnée avec la glutathione S-transférase (GST). Un vecteur d'expression désigné dans ce cas est représenté par le vecteur plasmidique pGEX-4T-3 (Pharmacia ref-27.4583).

15 L'invention vise en outre les cellules hôtes transfectées par ces vecteurs précédents. Ces cellules peuvent être obtenues par l'introduction dans des cellules hôtes d'une séquence nucléotidique insérée dans un vecteur tel que défini ci-dessus, puis la mise en culture desdites cellules dans des conditions permettant la réplication et/ou l'expression de la séquence nucléotidique transfectée.

20 Ces cellules sont utilisables dans une méthode de production d'un polypeptide recombinant de séquence SEQ ID n° 2, SEQ ID n° 4, SEQ ID n° 6, SEQ ID n° 8 ou SEQ ID n° 10, ou tout fragment ou dérivé biologiquement actif de celui-ci.

25 La méthode de production d'un polypeptide de l'invention sous forme recombinante est elle-même comprise dans la présente invention, et se caractérise en ce que l'on cultive les cellules transfectées dans des conditions permettant l'expression d'un polypeptide recombinant de séquence SEQ ID n° 2, SEQ ID n° 4, SEQ ID n° 6, SEQ ID n° 8 ou SEQ ID n° 10, ou de tout fragment
30 ou dérivé biologiquement actif de celui-ci, et que l'on récupère ledit polypeptide recombinant.

35 Les procédés de purification utilisés sont connus de l'homme du métier. Le polypeptide recombinant peut être purifié à partir de lysats et extraits cellulaires, du surnageant du milieu de culture, par des méthodes utilisées individuellement ou en combinaison, telles que le fractionnement, les méthodes de chromato-

graphie, les techniques d'immunoaffinité à l'aide d'anticorps mono ou polyclonaux spécifiques, etc.

Une variante préférée consiste à produire un polypeptide recombinant fusionné à une protéine "porteuse" (protéine chimère). L'avantage de ce système est qu'il permet une stabilisation et une diminution de la protéolyse du produit recombinant, une augmentation de la solubilité au cours de la renaturation *in vitro* et/ou une simplification de la purification lorsque le partenaire de fusion possède une affinité pour un ligand spécifique.

Avantageusement, les polypeptides de l'invention sont fusionnés avec la glutathione S-transférase en position N-terminale (système "GST" Pharmacia). Le produit de fusion est dans ce cas détecté et quantifié grâce à l'activité enzymatique de la GST. Le réactif colorimétrique utilisé est un accepteur de glutathion, substrat de la GST. Le produit recombinant est purifié sur un support de chromatographie auquel ont été préalablement couplées des molécules de glutathion.

Les anticorps mono ou polyclonaux capables de reconnaître spécifiquement une protéine SR-p70 selon la définition donnée précédemment font également partie de l'invention. Des anticorps polyclonaux peuvent être obtenus à partir du sérum d'un animal immunisé contre la protéine, produite par exemple par recombinaison génétique suivant la méthode décrite ci-dessus, selon les modes opératoires usuels.

Les anticorps monoclonaux peuvent être obtenus selon la méthode classique de culture d'hybridomes décrite par Köhler et Milstein, Nature, 1975, 256, 495-497.

Dès anticorps avantageux sont des anticorps dirigés contre la région centrale, résidu 110-310.

Les anticorps selon l'invention sont par exemple des anticorps chimériques, des anticorps humanisés, des fragments Fab et F(ab')₂. Ils peuvent également se présenter sous forme d'immunoconjugués ou d'anticorps marqués.

Par ailleurs, outre leur utilisation pour la purification des polypeptides recombinants, les anticorps de l'invention, en particulier les anticorps monoclonaux, peuvent également être utilisés pour la détection de ces polypeptides dans un échantillon biologique.

Ils constituent ainsi un moyen d'analyse immunocytochimique ou immunohistochimique de l'expression de protéines SR-p70 sur des coupes de

tissus spécifiques, par exemple par immunofluorescence, marquage à l'or, immunoperoxydase, ...

Ils permettent notamment de mettre en évidence une accumulation anormale de protéines SR-p70 dans certains tissus ou prélèvements biologiques, ce qui les rend utiles pour la détection des cancers ou le suivi de l'évolution ou de la rémission de cancers préexistants.

Plus généralement, les anticorps de l'invention peuvent être avantageusement mis en oeuvre dans toute situation où l'expression d'une protéine SR-p70 doit être observée.

L'invention concerne donc également un procédé de diagnostic *in vitro* de pathologies corrélées à une expression ou une accumulation anormale de protéines SR-p70, notamment les phénomènes de cancérisation, à partir d'un prélèvement biologique, caractérisé en ce que l'on met en contact au moins un anticorps de l'invention avec ledit prélèvement biologique, dans des conditions permettant la formation éventuelle de complexes immunologiques spécifiques entre une protéine SR-p70 et le ou lesdits anticorps et en ce que l'on détecte les complexes immunologiques spécifiques éventuellement formés.

L'invention concerne également un kit pour le diagnostic *in vitro* d'une expression ou une accumulation anormale de protéines SR-p70 dans un prélèvement biologique et/ou pour la mesure du taux d'expression de celle-ci dans ledit prélèvement comprenant :

- au moins un anticorps spécifique d'une protéine SR-p70, éventuellement fixé sur un support,

- des moyens de révélation de la formation de complexes antigènes/anticorps spécifiques entre une protéine SR-p70 et ledit anticorps et/ou des moyens de quantification de ces complexes.

L'invention vise également une méthode de diagnostic précoce de la formation des tumeurs, par la mise en évidence dans le sérum d'un individu, d'auto-anticorps dirigés contre une protéine SR-p70.

Une telle méthode de diagnostic précoce est caractérisé en ce que l'on met en contact un échantillon de sérum prélevé chez un individu avec un polypeptide de l'invention, éventuellement fixé sur un support, dans des conditions permettant la formation de complexes immunologiques spécifiques entre ledit polypeptide et les auto-anticorps éventuellement présents dans l'échantillon de sérum, et en ce que l'on détecte les complexes immunologiques spécifiques éventuellement formés.

L'invention comprend également des compositions pharmaceutiques comprenant comme principe actif un polypeptide répondant aux définitions précédentes, préférentiellement sous forme soluble, associé à un véhicule pharmaceutiquement acceptable.

5 De telles compositions offrent une nouvelle approche pour traiter les phénomènes de cancérisation au niveau du contrôle de la multiplication et la différenciation cellulaire.

Préférentiellement, ces compositions peuvent être administrées par voie systématique, de préférence **par voie** intraveineuse, par voie intramusculaire, intradermique ou par voie orale.

10 Leurs modes d'administration, posologies et formes galéniques optimaux peuvent être déterminés selon les critères généralement pris en compte dans l'établissement d'un traitement thérapeutique adapté à un patient comme par exemple l'âge ou le **poids** corporel du patient, la gravité de son état général, la tolérance au traitement et les effets secondaires constatés, etc.

15 L'invention comprend enfin une méthode de thérapie génique dans laquelle des séquences nucléotidiques codant pour une protéine SR-p70 sont transférées à des cellules cibles par le biais de vecteurs viraux inactivés.

20 D'autres caractéristiques et avantages de l'invention apparaissent dans la suite de la description avec les exemples et les figures dont les légendes sont représentées ci-après.

25 **LEGENDE DES FIGURES**

25

Figure 1 : Comparaison nucléique de l'ADNc du SR-p70a de singe (correspondant à SEQ ID n° 1) avec la séquence nucléique de l'ADNc de p53 de singe.

30

Figure 2 : Comparaison protéique de SR-p70a de singe avec la protéine p53 de singe (sw : p53-cerae).

35

Figure 3 : Comparaison de la séquence nucléique de l'ADNc de SR-p70a et b de singe (correspondant respectivement à SEQ ID n° 1 et SEQ ID n° 3).

Figure 4 : Séquence nucléique et séquence protéique déduite de SR-p70a de singe.

Figure 5 : Séquence nucléique partielle et séquence protéique déduite complète de SR-p70b de singe.

Figure 6 : Séquence nucléique partielle et séquence protéique complète déduite de SR-p70 d'homme (correspondant à SEQ ID n° 5).

Figure 7 : Séquence nucléique partielle et séquence protéique déduite complète de SR-p70c de souris (correspondant à SEQ ID n° 7).

Figure 8 : Séquence nucléique partielle et séquence protéique déduite partielle de SR-p70a de souris (correspondant à SEQ ID n° 9).

Figure 9 : Multialignement des protéines déduites des ADNc SR-p70 de singe (a et b), d'homme et de souris (a et c).

Figure 10a : Immunoempreinte de la protéine SR-p70.

Figure 10b : Détection de la protéine endogène SR-p70.

Figure 11 : Localisation chromosomique du gène SR-p70 humain. Le signal apparaît sur le chromosome 1, dans la région p36.

EXEMPLE I

Clonage de l'ADNc du SR-p70 de cellules COS-3.

1. Culture des cellules COS-3

Les cellules COS-3 (cellules de rein de singe vert d'Afrique transformées par l'antigène T du virus SV 40) sont cultivées dans le milieu DMEM (GIBCO-BRL) référence 41 965-047) contenant 2 mM de L-glutamine et supplémenté avec 50 mg/l de gentamycine et de 5 % de sérum de bovin foetal (GIBCO-BRL référence 10231-074) jusqu'à semi-confluence.

2. Préparation de l'ARN messenger

a) extraction de l'ARN messenger

Les cellules sont récupérées de la façon suivante :

5 - les cellules adhérentes sont lavées deux fois avec du tampon PBS (phosphate buffered saline, référence 04104040-GIBCO-BRL) puis grattées avec un grattoir en caoutchouc et centrifugées.

10 Le culot cellulaire est mis en suspension dans le tampon de lyse de composition suivante : guanidine-thiocyanate 4M ; citrate de sodium 25mM pH 7 ; sarcosyl 0,5 % ; β -mercaptoéthanol 0,1 M. La suspension est soniquée à l'aide d'un sonicateur ultra turax n° 231256(Janke et Kundel) à puissance maximale pendant une minute. On ajoute de l'acétate de sodium pH 4 jusqu'à 0,2 M. La solution est extraite avec un volume d'un mélange phénol/chloroforme (v/v ; 5/1). On précipite à -20°C l'ARN contenu dans la

15 phase aqueuse à l'aide d'un volume d'isopropanol. Le culot est resuspendu dans le tampon de lyse. La solution est à nouveau extraite avec un mélange phénol/chloroforme et l'ARN est précipité avec de l'isopropanol. Après lavage du culot avec de l'éthanol 70 % puis 100 %, l'ARN est resuspendu dans de l'eau.

b) Purification de la fraction poly A* de l'ARN

20 La purification de la fraction poly A* de l'ARN est réalisée à l'aide du kit Dynabeads oligo (dT)₂₅ de DYNAL (référence 610.05) suivant le protocole préconisé par le fabricant. Le principe est basé sur l'utilisation de billes polystyrène super-paramagnétique sur lesquelles est attaché un

25 oligonucléotide poly(dT)₂₅. La fraction poly A* de l'ARN est hybridée sur l'oligo(dT)₂₅ couplé aux billes que l'on piège sur un support magnétique.

3. Constitution de la banque d'ADN complémentaire

a) préparation de l'ADN complémentaire

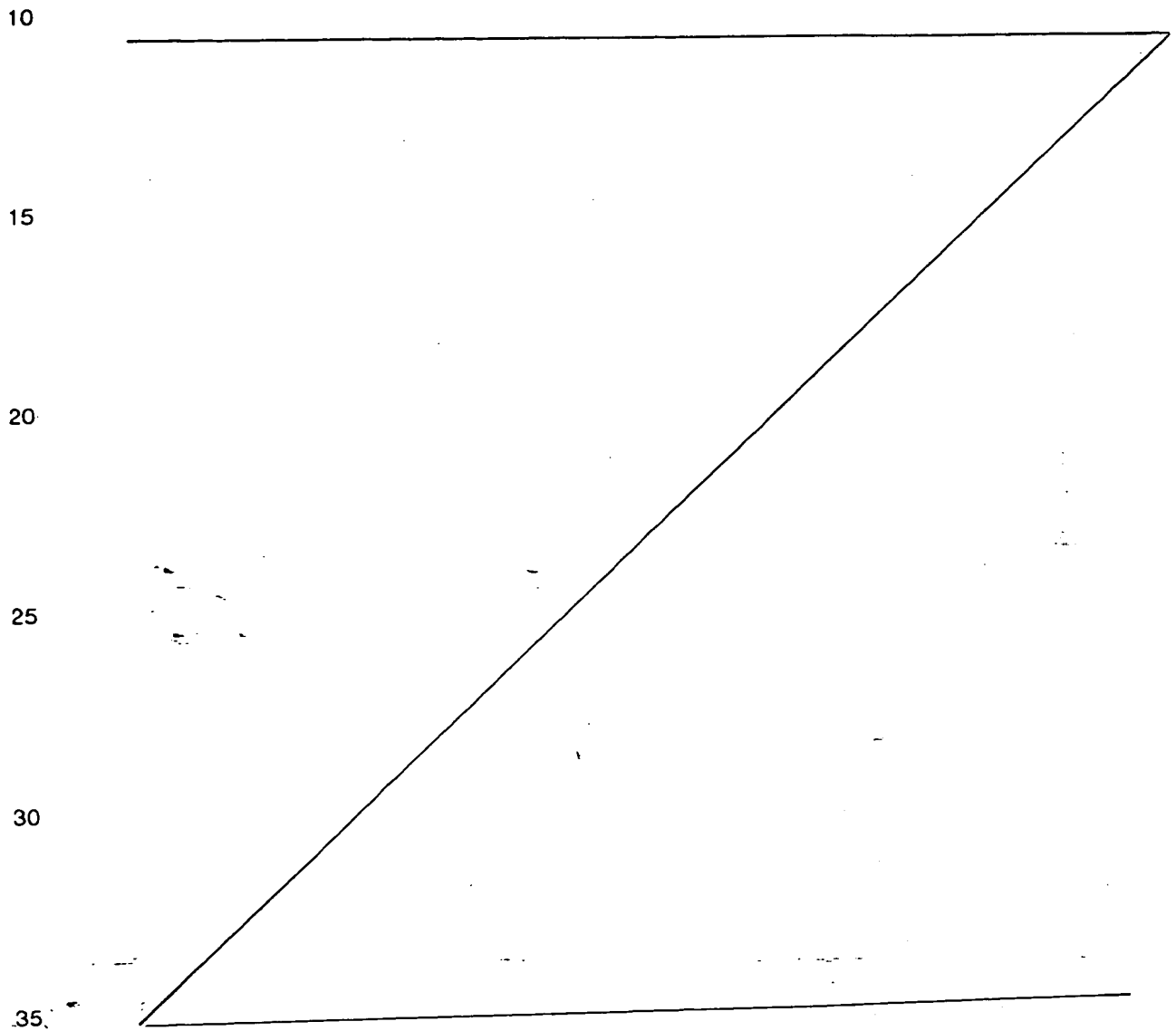
30 A partir de 0,5 μg des ARN-poly A* de cellules COS-3 obtenues à l'issue de l'étape 2, on prépare l'ADN complémentaire simple-brin marqué au ^{32}P .dCTP (l'ADN complémentaire obtenu présente une activité spécifique de 3000 dpm/ng) avec l'amorce synthétique de séquence suivante (comprenant un site BamHI) :

35 5'<GATCCGGGCC CTTTTTTTTT TTT<3'

Dans un volume de 30 μ l de tampon de composition : Tris HCl 50 mM pH 8,3, $MgCl_2$ 6 mM, DTT 10 mM, KCl 40 mM, contenant 0,5 mM de chacun des désoxynucléotides triphosphates, 30 μ Ci de dCTP $\alpha^{32}P$ et 30 U de RNasin (promega). Après une heure d'incubation à 37°C, puis 10 minutes à 50°C, puis de nouveau 10 minutes à 37°C, avec 200 unités de l'enzyme transcriptase inverse RNase H⁻ (GIBCO-BRL référence 8064A), on ajoute 4 μ l d'EDTA.

b) Hydrolyse alcaline de la matrice ARN

On ajoute 6 μ l d'une solution de NaOH 2N, puis on incube pendant 5 minutes à 65°C.



c) Purification sur colonne sephacryl S400

Afin d'éliminer l'amorce synthétique, on purifie l'ADN complémentaire sur une colonne de 1 ml de sephacryl S400 (Pharmacia), équilibrée dans du tampon TE.

5 Les deux premières fractions radioactives sont regroupées et précipitées avec 1/10 de volume d'une solution d'acétate d'ammonium 10 M et 2,5 volumes d'éthanol, ceci après une extraction, avec un volume de chloroforme.

d) Addition homopolymérique de dG

10 On allonge l'ADN complémentaire en 3' avec une "queue" de dG avec 20 unités de l'enzyme terminale transférase (Pharmacia 27073001). On incube dans 20 µl de tampon de composition : Tris HCl 30 mM pH 7,6 ; chlorure de cobalt 1mM, acide cacodylique 140 mM, DTT 0,1mM, dGTP 1 mM, pendant 15 minutes à 37°C, puis on ajoute 2 µl d'EDTA 0,5 M.

e) On répète à nouveau les étapes b) et c)

15 f) Appariement du vecteur de clonage pSE1 (EP 506 574) et de l'ADN complémentaire en présence de l'adaptateur.

On centrifuge, le culot est dissous dans 33 µl de tampon TE, on ajoute 5 µl (125 ng) de vecteur de clonage pSE1, 1 µl(120 ng) de l'adaptateur de séquence suivante (comprenant un site Apal),

20 5'AAAAAAAAAAAAAGGGCCCG3'

10 µl d'une solution de NaCl 200 mM, on incube pendant 5 minutes à 65°C puis on laisse refroidir le mélange réactionnel jusqu'à température ambiante.

g) Ligation

25 On ligue le vecteur de clonage et l'ADNc simple brin dans un volume de 100 µl avec 32,5 unités de l'enzyme ADN ligase du phage T4 (Pharmacia référence 270 87002) pendant une nuit à 15°C dans un tampon de composition : Tris HCl 50 mM pH 7,5, MgCl₂ 10 mM, ATP 1 mM.

h) Synthèse du deuxième brin de l'ADNc

30 On élimine les protéines par extraction au phénol suivie d'une extraction au chloroforme, puis on ajoute 1/10ème de volume d'une solution d'acétate d'ammonium 10 mM, puis 2,5 volumes d'éthanol. On centrifuge, le culot est dissous dans le tampon de composition Tris acétate 33 mM pH 7,9 acétate de potassium 62,5 mM, acétate de magnésium 1 mM et dithiothréitol (DTT) 1 mM, le deuxième brin d'ADN complémentaire est synthétisé dans un volume de 30 µl avec 30 unités de l'enzyme ADN polymérase du phage T4 (Pharmacia, référence 270718) et un mélange de 1 mM des quatre désoxynucléotides

35

triphosphates de dATP, dCTP, dGTP et dTTP, ainsi que deux unités de la protéine du gène 32 du phage T4 (Pharmacia, référence 27-0213)) pendant une heure à 37°C. On extrait au phénol et on retire les traces de phénol par une colonne de polyacrylamide P10 (Biogel P10-200-400 mesh - référence 15011050 - Biorad).

i) Transformation par électroporation

On transforme des cellules *E. Coli* MC 1061 avec l'ADN recombinant obtenu précédemment par électroporation à l'aide de l'appareil Biorad Gene Pulser (Biorad) utilisé à 2,5 kV dans les conditions prescrites par le fabricant, puis on fait pousser les bactéries pendant une heure dans du milieu dit milieu LB (Sambrook *op cite*) de composition : bactotryptone 10 g/l ; extrait de levure 5 g/l ; NaCl 10 g/l.

On détermine le nombre de clones indépendants en étalant une dilution au 1/1000ème de la transformation après la première heure d'incubation sur une boîte de milieu LB additionné de 1,5 % d'agar (p/v) et de 100 µg/ml d'ampicilline, appelé par la suite milieu LB gélosé. Le nombre de clones indépendants est de 1 million.

j) Analyse des ADNc de la banque

Dans le cadre de l'analyse de clones individualisés de la banque par un séquençage nucléique du 5' des ADNc, un clone, dénommé SR-p70a s'est révélé présenter une homologie partielle avec l'ADNc de la protéine déjà connue, la protéine p53 (Genbank X 02469 et X 16384) (Figure 1). Les séquences ont été réalisées avec le kit United States Biochemical (référence 70770) et/ou le kit Applied Biosystems (références 401434 et/ou 401628) qui utilisent la méthode de Sanger et al., Proc. Natl. Acad. Sci. USA ; 1977, 14, 5463-5467. L'ADN plasmidique est préparé à partir du kit WIZARD mini préparation (Promega référence A7510). Les amorces utilisées sont des oligonucléotides de 16 à 22 mer, complémentaires soit au vecteur pSE1 dans la région immédiatement en 5' de l'ADNc, soit à la séquence de l'ADNc.

Un second ADNc a été isolé à partir de la même banque en criblant de manière similaire à la technique décrite dans l'EXEMPLE III 3) ci-après avec un fragment de l'ADN SR-p70a a marqué au ³²P avec le kit BRL "Random Primers DNA labelling systems" (référence 18187-013). Le tampon d'hybridation et de lavage sont additionnés de 50 % de formamide. Le dernier lavage est réalisé en 0,1 x SSC/SDS 0,1 % à 60°C. Cette seconde séquence

(ADNc SR-p70b) est identique à la première mais présente un fragment interne délété (Figure 3).

Les deux ADNc SR-p70, d'une longueur de 2874 nucléotides (SR-p70a) et de 2780 nucléotides (SR-p70b) correspondent aux produits d'un seul gène, un épissage alternatif entraînant une délétion de 94 bases entre les nucléotides 1637 et 1732 et une terminaison prématurée de la protéine codée correspondante. Les protéines déduites des deux ADNc présentent respectivement 637 acides aminés et 499 acides aminés (Figures 4 et 5).

EXEMPLE II

Obtention de la séquence et clonage de l'ADNc de la protéine SR-p70 à partir de cellules HT-29 (Adénocarcinome de colon humain).

1) Culture des cellules HT-29

Les cellules sont cultivées en milieu Mac Coy 5 (GIBCO 26600-023) additionné de 10 % de sérum foetal de veau (GIBCO 10081-23) et 50 mg/l de gentamycine jusqu'à semi-confluence.

2) Préparation de l'ADN complémentaire

L'ARN messager est préparé comme décrit dans l'EXEMPLE I.2. L'ADNc est préparé de manière similaire à celle décrite dans l'EXEMPLE I.3 avec 5 µg d'ARN messager total en utilisant une amorce poly(T)₁₂. La réaction n'est pas interrompue avec de l'EDTA.

3) Amplification spécifique de l'ADNc humain par la technique dite de PCR

La polymérisation est réalisée avec 4 µl d'ADNc dans 50 µl final avec le tampon de composition suivante : Tris HCl 10 mM pH 8,3, MgCl₂ 2,5 mM, KCl 50 mM en présence de 10 % DMSO, dNTP 0,5 mM, 4 µg/ml de chacune des deux amorces nucléiques et de 2,5 unités de TAQ ADN polymérase (Boehringer). Les couples d'amorces ont été choisis sur la séquence nucléique du clone SR-p70 de COS-3, notamment en amont de l'ATG d'initiation et en aval du TGA d'arrêt de traduction et sont de compositions suivantes :

amorce sens : ACT GGT ACC GCG AGC TGC CCT CGG AG

site de restriction Kpn I

amorce antisens : GAC TCT AGA GGT TCT GCA GGT GAC TCA G
site de restriction Xba I

5

La réaction est réalisée durant 30 cycles 94°C/1 minute, 54-60°C/1 minute 30 secondes et 72°C/1 minute 30 secondes, suivi d'un dernier cycle de 72°C/6 minutes.

10

4) *Obtention de la séquence de l'ADNc humain*

Dans un premier temps, le produit de PCR est éliminé des oligonucléotides sur une colonne de sephacryl S400 puis déssalé par chromatographie d'exclusion sur une colonne de polyacrylamide P10 (Biorad référence 1504144). Les réactions de séquençage sont réalisées à l'aide du kit Applied Biosystems (référence 401628) avec des oligonucléotides spécifiques de l'ADNc. La séquence obtenue est très similaire à celle du SR-p70a de singe et la protéine déduite contient 636 acides aminés (Figure 6).

15

De manière similaire, d'autres séquences issues de lignées ou de tissus humain ont été obtenues pour la partie codante du SR-p70 humain, notamment à partir du poumon ou du pancréas. Les protéines déduites de ces séquences sont identiques à celles obtenues pour la lignée HT-29.

20

5) *Clonage de l'ADNc humain dans le plasmide pCDNA3 (Pharmacia)*

Le produit PCR obtenu en 3) ainsi que le plasmide sont digérés par les deux enzymes de restriction Kpn I et Xba I puis purifiés après migration sur un gel d'agarose 1 % à l'aide du kit geen clean (Bio 101 référence 3105). Après ligation avec 100 ng d'insert et 10 ng de vecteur et transformation (technique décrite dans l'EXEMPLE 1 3)g) et i), les clones recombinants sont vérifiés par séquençage à l'aide du kit Applied Biosystems cité ci-dessus.

25

30

35

EXEMPLE III

Clonage de l'ADNc du SR-p70 de souris à partir de cellules AtT-20 (tumeur hypophysaire)

5

1) Culture cellulaire de la lignée AtT-20

Les cellules sont cultivées dans du milieu Ham F10 (GIBCO 31550-023) additionné de 15 % de sérum de cheval (GIBCO 26050-047) et de 2,5 % de sérum foetal de veau (GIBCO 10081-073) et de 50 mg/l de gentamycine jusqu'à semi-confluence.

10

2) Préparation de la banque d'ADN complémentaire

La banque est réalisée comme décrit dans l'EXEMPLE I, II et III à partir des cellules cultivées ci-dessus.

15

3) Criblage de la banque

a) Préparation des membranes

Les clones de la banque sont étalés sur du milieu LB gélosé (boîtes de petri diamètre 150) revêtu de membranes Biodyne A (PALL référence BNNG 132). Après une nuit à 37°C, les clones sont transférés par contact sur de nouvelles membranes. Ces dernières sont traitées en les déposant sur du papier Wathman 3 MM imbibé des solutions suivantes : NaOH 0.5 N, NaCl 1.5 M pendant 5 minutes puis Tris HCl 0.5 M pH 8, NaCl 1.5 M pendant 5 minutes. Après un traitement à la protéinase K dans le tampon suivant : Tris HCl 10 mM pH 8, EDTA 10 mM, NaCl 50 mM, SDS 0.1 %, protéinase K 100 µg/ml pendant une heure à température ambiante, les membranes sont lavées abondamment dans du 2 x SSC (sodium citrate NaCl), séchées, puis incubées au four sous vide à 80°C pendant 20 minutes.

20

25

b) Préparation de la sonde

30

Sur la base de séquences des ADNc SR-p70 de singe et d'humain, une première séquence a été réalisée sur un fragment amplifié à partir de l'ARNm de la lignée AtT-20 comme décrit dans l'EXEMPLE II 3) et 4) avec les oligomères de compositions suivantes :

35

amorce sens : GCC ATG CCT GTC TAC AAG
amorce antisens : ACC AGC TGG TTG ACG GAG

Sur la base de cette séquence, une sonde oligomérique spécifique de souris a été choisie et présente la composition suivante :

GAG CAT GTG ACC GAC ATT G.

5 100 ng de la sonde sont marqués en 3' avec 10 unités de Terminal Transférase (Pharmacia) et 100 μ Ci de dCTP α^{32} P 3000 Ci/mme (Amersham référence PB 10205) dans 10 μ l du tampon suivant : Tris HCl 30 mM pH 7.6 , acide cacodylique 140 mM, CoCl₂ 1 mM, DTT 0.1 mM pendant 15 minutes à 37°C. Les nucléotides radiomarqués non incorporés sont éliminés sur une
10 colonne de polyacrylamide P10 (Biorad, référence 1504144). La sonde obtenue a une activité spécifique environ de 5.10⁸ dpm/ μ g.

c) Préhybridation et hybridation

Les membranes préparées en a) sont préhybridées 30 minutes à 42°C dans 6 x SSC, 5 x Denhart's, 0,1 % SDS puis hybridées quelques heures dans le
15 même tampon additionné de la sonde préparée en b) à raison de 10⁶ dpm/ml.

d) Lavage et exposition des membranes

Les membranes sont lavées deux fois à température ambiante dans le tampon 2 x SSC/SDS 0.1 % puis une heure à 56°C en 6 x SSC/SDS 0.1 %. Les clones hybridés sont révélés avec des films KODAK XOMAT. Un clone positif contenant le SR-p70 de souris est sélectionné et dénommé ci-après
20 SR-p70c.

4) Séquençage du SR-p70 de souris et analyse de la séquence

25 La séquence est obtenue à l'aide du kit Applied Biosystem (référence 401628). La séquence protéique déduite de l'ADNc SR-p70c de souris (Figure 7) présente une très forte homologie avec celles de l'humain et de singe excepté dans la partie N terminale qui diverge fortement (voir Figure 9). A l'aide de la technique dite de PCR, de manière similaire à celle décrite dans l'EXEMPLE II, 3) et 4), une seconde séquence 5' (issue de la même banque
30 AtT-20) a été obtenue (Figure 8). La séquence protéique N terminale déduite (séquence dénommé SR-p70a) est très similaire à celle déduite des ADNc SR-p70 humain et de singe (SR-p70a) (Figure 9). La lignée AtT-20 présente donc au moins deux transcripts SR-p70. Ces 2 derniers divergent dans la partie N terminale par des épissages différents.

EXEMPLE IV

1) Production de protéine recombinante SR-p70 dans *E. coli*

5 a) Construction du plasmide d'expression

Elle consiste à mettre la partie -COOH terminale de la protéine SR-p70a de singe, depuis la valine en position 427 à l'histidine -COOH terminale en position 637, en fusion avec la glutathione S-transferase (GST) du vecteur plasmidique pGEX-4T-3 (Pharmacia référence 27-4583). Pour cela, l'insert correspondant de la SR-p70a (position 1434 à 2066) a été amplifié par PCR avec 10 ng de plasmide contenant l'ADNc SR-p70a de singe. Les amorces nucléiques sont de composition suivante :

15 amorce sens : TTT GGA TCC GTC AAC CAG CTG GTG GGC CAG
site de restriction BamHI

amorce antisens : AAA GTC GAC GTG GAT CTC GGC CTC C
site Sal I

20 Le fragment obtenu ainsi que le vecteur sont digérés par les enzymes de restriction BamHI et Sal I et le clonage est réalisé comme décrit dans l'EXEMPLE II 5). Le clone sélectionné est appelé pG SR-p70.

b) Expression et purification de la protéine fusion GST-pSR-p70

25 Cette étape a été réalisée en utilisant le kit "bulk GST purification module" (Pharmacia Référence 27-4570-01).

De manière schématique, le clone recombinant a été mis en culture à 37°C dans un litre de milieu 2x YTA + ampicilline 100 µg/ml. A DO 0,8, l'expression est induite avec 0,5 mM d'IPTG pendant 2 heures à 37°C. Après centrifugation, le culot cellulaire est repris dans du PBS froid puis soniqué par ultrason. Après adjonction de 1 % triton X-100, on incube 30 minutes sous agitation à température ambiante. Après centrifugation à 12 000 g, 10 minutes à 4°C, on récupère le surnageant. La purification est ensuite réalisée sur une colonne de chromatographie d'affinité glutathion sepharose 4B. La fixation et le lavage sont réalisées en tampon PBS et l'élution est réalisée par compétition avec du glutathion réduit. La concentration finale est amenée à 300 µg/ml de protéine fusion.

2) Production de protéine SR-p70a dans les cellules COS-3.

Les cellules COS-3 sont transfectées avec de l'ADN plasmidique pSE1 dans lequel a été cloné l'ADNc de SR-p70a de singe (EXEMPLE I 1)) ou avec
 5 de l'ADN plasmidique du vecteur PSE1 en tant que témoin par la technique du DEAE Dextran : les cellules COS-3 sontensemencées à 5×10^5 cellules par boîte de 6 cm en milieu de culture contenant 5 % de sérum de bovin foetal (EXEMPLE I 1)). Après culture, les cellules sont rincées avec du PBS. On
 10 ajoute 1 ml du mélange suivant : milieu contenant $6,5 \mu\text{g}$ d'ADN, $250 \mu\text{g/ml}$ de DEAE Dextran et $100 \mu\text{M}$ de chloroquine. Les cellules sont incubées à 37°C en 5 % CO_2 durant 4 à 5 heures. Le milieu est aspiré, on ajoute 2 ml de PBS additionné de 10 % DMSO et les cellules sont incubées pendant une minute en
 15 remuant légèrement les boîtes. Le milieu est à nouveau aspiré et les cellules sont rincées deux fois avec du PBS. Les cellules sont alors incubées à 37°C avec du milieu contenant 2 % de sérum de bovin foetal pendant la durée de l'expression qui est généralement de 3 jours.

La protéine SR-p70a est alors analysée comme décrit dans l'EXEMPLE VI
 20 par immunoempreinte.

EXEMPLE V

Préparation d'anticorps spécifiques

25 150 μg de protéines de l'échantillon préparé selon l'EXEMPLE IV ont été utilisés pour immuniser un lapin (mâle de 1,5 à 2 kg environ, New-Zealand). Les immunisations ont été effectuées tous les 15 jours selon le protocole décrit par Vaitukaitis, Methods in Enzymology, 1981, 73, 46. Pour la première
 30 injection, un volume de solution antigénique est émulsifié par un volume d'adjuvant complet de Freund (Sigma référence 4258). Cinq rappels ont été administrés en adjuvant incomplet de Freund (Sigma référence 5506).

EXEMPLE VI

35 Détection de la protéine SR-p70 "Western immunoblotting" (immunoempreinte)

1) Matériels utilisés pour l'immunoempreinte

a) Lignées cellulaires utilisées pour l'immunoempreinte.

Les lignées cellulaires suivantes ont été cultivées, comme décrit dans le catalogue "catalogue of cell lines and hybridomas, 7th edition, 1992 de l'ATCC (American Type Culture Collection) : COS-3, CV-1 (lignée de cellules de rein de singe), HT-29, U-373MG (glioblastome humain), MCF7 (adénocarcinome mammaire humain), SKNAS (neuroblastome humain cultivé dans les mêmes conditions que COS-3), **SK-N-MC** (neuroblastome humain), IMR-32 (neuroblastome humain), **CHP212** (neuroblastome humain cultivé dans les mêmes conditions que **CV-1**), Saos-2 (ostéosarcome), SK-OV-3 (adénocarcinome d'ovaire) et **SW 480** (adénocarcinome de colon humain).

b) Cellules COS-3 transfectées par l'ADNc SR-p70a.

Les cellules COS-3 ont été transfectées comme décrit dans l'EXEMPLE IV 2). En tant que témoin, les cellules ont été transfectées avec de l'ADN plasmidique PSE1 ne contenant pas l'ADNc recombinant SR-p70a.

2) Préparation des échantillons protéiques à partir de culture cellulaire eucaryote ou de cellules transfectées.

Après culture, les cellules sont lavées avec du PBS puis reprises dans un tampon RIPA (PBS avec 1 % NP40, 0,5 % sodium déoxycholate, 0,5 % SDS) complémenté avec 10 µg/ml RNase A, 20µg/ml DNase 1, 2 µg/ml aprotinine, 0,5 µg/ml leupeptine, 0,7 µg/ml pepstatine et 170 µg/ml PMSF. Les cellules sont soniquées par ultrason à 4°C et laissées 30 minutes à 4°C. Après microcentrifugation à 12 000 rpm, on récupère le surnageant. La concentration de protéine est mesurée par la méthode de Bradford.

3) "Western blotting"

5 ou 50 µg de protéines (50 µg pour les lignées cellulaires et 5 µg pour des cellules transfectées) sont mis dans 0,2 volume du tampon d'électrophorèse 6X suivant : Tris HCl 0,35 mM pH 6,8 10,3 % SDS, 36 % glycérol, 0,6 mM DTT, 0,012 % bleu de bromophénol. Les échantillons sont déposés et mis à migrer dans un gel SDS-Page 10 % (30/0,8 Bis) puis électrotransférés sur une membrane de nitrocellulose.

4) Révélation par l'anticorps

La membrane est incubée 30 minutes dans le tampon de blocage TBST (Tris HCl 10 mM pH 8, NaCl 150 mM, 0,2 % Tween 20) additionné de 5 % de lait (GIBCO- SKIM MILK) à température ambiante. La membrane est successivement mise en présence de l'anticorps anti-SR-p70 (α SR-p70) dans le même tampon 16 heures à 4°C, lavée 3 fois pendant 10 minutes avec du TBST, puis incubée une heure à 37°C avec un second anticorps anti-immunoglobuline de lapin couplé avec de la peroxydase (SIGMA A055). Après trois lavages de 15 minutes, la révélation est effectuée en utilisant le kit ECL (Amersham RPN2106) par chimioluminescence.

Parallèlement, les mêmes échantillons ont été révélés par un anticorps anti-P53 (α p53) (sigma BP5312) suivi d'un second anticorps anti-immunoglobuline de souris.

5) Figure et résultat.

Figure 10 : Immunoempreinte de la protéine SR-p70

Figure 10a : Détection de la protéine recombinante SR-p70

- colonnes 1 et 3 : COS-3 transfectée par le vecteur pSE-1.

- colonnes 2 et 4 : COS-3 transfectée par le plasmide pSE1 contenant l'ADNc du SR-p70a.

- colonnes 1 et 2 : Révélation par l'anticorps anti-SR-p70 (α SR-p70).

- colonnes 3 et 4 : Révélation par l'anticorps anti-P53 (α p53).

Figure 10b : Détection de la protéine endogène SR-p70

- colonnes 1 : COS-3 ; 2 : CV-1 ; 3 : HT-29 ; 4 : U-373 MG ; 5 : MCF7 ; 6 : SKNAS ; 7 : SK-N-MC ; 8 : IMR-32 ; 9 : CHP212 ; 10 : Saos-2 ; 11 : SK-OV-3 et 12 : SW480.

A : Révélation par l'anticorps α SR-p70

B : Révélation par l'anticorps α p53

L'anticorps α SR-p70 reconnaît spécifiquement les protéines recombinantes (Figure 10a) et endogènes (Figure 10b) et ne croise pas avec la p53. L'analyse de lignées cellulaires humaines ou de singe montre que la protéine SR-p70 comme la p53 est généralement faiblement détectable. Par contre, lorsqu'une accumulation de p53 existe, la SR-p70 devient elle aussi plus facilement

déTECTABLE (Figure 10b). Une étude par RT-PCR de la distribution des transcrits SR-p70 montre que le gène est exprimé dans tous les types cellulaires testés.

EXEMPLE VII

5

Clonage du gène du SR-p70 et localisation chromosomique.

1) Clonage du gène SR-p70

La banque utilisée est une banque de cosmides, préparée avec de l'ADN génomique humain purifié de placenta, et commercialisée par Stratagène (référence 95 1202).

Le criblage du gène est réalisé comme décrit dans l'exemple III 3) avec un fragment d'ADN SR-p70 marqué au ^{32}p avec le kit BRL "Random Primers DNA Labelling Systems" (référence 18187-013). Les tampons d'hybridation et de lavage sont additionnés de 50 % formamide. Le dernier lavage est réalisé en 0,1 x SSC/SDS 0,1 % à 60°C. De manière similaire, le gène SR-p70 a été isolé à partir d'une banque préparée avec de l'ADN génomique de la souris black C57.

20

Une analyse et un séquençage partiel des clones mettent en évidence la présence d'un minimum de 12 exons avec une structure rappelant celle du gène p53.

2) Localisation chromosomique du gène SR-p70 chez l'homme

25

Elle a été réalisée avec de l'ADN du gène SR-70 humain en utilisant la technique décrite par R. Slim et al., Hum. Genet., 1991, 88, 21-26. Cinquante mitoses ont été analysées dont plus de 80% avaient des doubles spots localisés en 1p36 sur les deux chromosomes et plus particulièrement en 1p36.2-1p36.3 (Figure 11). L'identification du chromosome 1 et son orientation sont basées sur l'hétérochromatine de la constriction secondaire. Les images ont été faites sur un microscope Zeiss Axiophot, saisies par une caméra CCD refroidie LHESA et traitées par Optilab.

35

L'homologie de structure entre le domaine de fixation à l'ADN de la p53 et la région centrale de la protéine SR-p70 permet d'inférer que la SR-p70 est un facteur de transcription (cf. figure 1 et 2). En effet, la p53 (393 acides aminés) est constituée de plusieurs domaines fonctionnels. La région N-terminale (1-91 acides aminés) est impliquée dans l'activation de la transcription, et contient des sites d'interaction à différentes protéines cellulaires et virales. La partie centrale (acides aminés 92 à 292) permet la fixation aux séquences d'ADN spécifiques situées dans les régions promotrices de certains gènes (la majorité des mutations ponctuelles inactivant la p53 sont localisées dans cette région), elle présente également de nombreux sites d'interaction avec des protéines virales qui inhibent son activité. Enfin, les 100 derniers acides aminés de la p53 sont responsables de son oligomérisation ainsi que de la régulation de celle-ci (Hainaut P., *Current Opinion in Oncology*, 1995, 7, 76-82 ; Prokocimer M., *Blood*, 1994, 84 n°8, 2391-2411).

L'homologie de séquence entre P53 et SR-p70 est significative notamment en ce qui concerne les acides aminés impliqués directement dans l'interaction à l'ADN suggérant que la SR-p70 se fixe aux sites p53 sur l'ADN. Ces acides aminés correspondent très exactement à ce qu'on appelle les "hot spot", acides aminés fréquemment mutés dans les tumeurs humaines (SWISS PROT : SW : P53_human et Prokocimer M., *Blood*, 1994, 84 n°8, 2391-2411). De cette homologie, on peut déduire que la protéine SR-p70 exerce un contrôle sur l'activité des gènes régulés par la p53, soit indépendamment de celle-ci soit en formant des hétérooligomères avec cette dernière.

En conséquence, à l'instar de la p53, les produits du gène SR-p70 doivent être impliqués dans le contrôle et la régulation du cycle cellulaire provoquant des arrêts du cycle (momentanés ou définitifs), et la mise en oeuvre de programmes tels que : la réparation de l'ADN, la différenciation ou la mort cellulaire. L'existence d'activités "p53-like" avait été fortement pressentie avec la mise en évidence chez les souris p53^{-/-}, d'activités de réparation de l'ADN et de mort cellulaire en réponse aux radiations ionisantes (Strasser. et al., *Cell*, 1994, 79, 329-339). Les auteurs de la présente invention ont localisé le gène SR-p70 humain dans la région télomérique du bras court du chromosome 1, précisément en 1p36.2-36.3, la plus petite région délétée (SRO) commune à une majorité de neuroblastomes et d'autres types de tumeurs (mélanomes et carcinomes) (White et al., *PNAS*, 1995, 92, 5520-5524). Cette région de perte d'hétérozygotie (LOH) délimite le locus d'un gène suppresseur de tumeur dont

la perte d'activité serait la cause de la formation des tumeurs. Il est important de rappeler que cette région est également sujette à "l'empreinte maternelle" ; l'allèle maternel est préférentiellement perdu dans les neuroblastomes présentant la délétion 1p36 (sans amplification de N-Myc) (Caron et al., Hum. Mol. Gen., 1995, 4, 535-539). Le gène sauvage SR-p70 introduit et exprimé dans des cellules de neuroblastome permet la réversion de leur transformation. La perte de cette activité anti-oncogénique est donc associée au développement de la tumeur. La région 1p36 présente une homologie synténique avec le segment distal du chromosome 4 de souris. Dans cette région a été localisée le gène *curly tail (ct)* (Beier et al., Mammalian Genome, 1995, 6, 269-272) impliqué dans les malformations congénitales du tube neural (MTN : *spida bifida*, anencéphalie...). La souris *ct* est le meilleur modèle animal d'étude de ces malformations. Il est admis que ces malformations résultent d'anomalies de la prolifération cellulaire. Compte tenu de la nature du gène SR-p70 et de sa localisation chromosomique, une des hypothèses est que le SR-p70 pourrait être l'homologue humain de *ct* et qu'à ce titre, la détection des mutations précoces et les anomalies chromosomiques concernant ce gène devraient permettre par exemple comme application, l'identification des personnes à risques (0,5-1 % des nouveaux-nés atteints par MTN), et la mise en oeuvre de traitements préventifs (Neumann et al., Nature Genetics, 1994, 6, 357-362 ; Di Vinci et al., Int. J. Cancer, 1994, 59, 422-426 ; Moll et al., PNAS, 1995, 92, 4407-4411 ; Chen et al., Development, 1995, 121, 681-691).

LISTE DES SEQUENCES

(1) INFORMATION GENERALE :

(i) DEPOSANT :

- (A) NOM : SANOFI
- (B) RUE : 32/34 rue MARBEUF
- (C) VILLE : PARIS
- (E) PAYS : FRANCE
- (F) CODE POSTAL : 75008

(ii) TITRE DE L'INVENTION : Protéine SR p70

(iii) NOMBRE DE SEQUENCES : 16

(iv) FORME LISIBLE PAR ORDINATEUR :

- (A) TYPE DE SUPPORT : Floppy disk
- (B) ORDINATEUR IBM PC compatible
- (C) SYSTEME D'EXPLOITATION : PC-DOS/MS-DOS
- (D) LOGICIEL : PatentIn Release #1.0, Version #1.25

(OEB)

(2) INFORMATION POUR LA SEQUENCE No 1 :

(i) CARACTERISTIQUES DE LA SEQUENCE :

- (A) LONGUEUR : 2874 paires de bases
- (B) TYPE : acide nucléique
- (C) NOMBRE DE BRINS : double
- (D) CONFIGURATION : linéaire

(ii) TYPE DE MOLECULE : ADNC

(vi) ORIGINE :

- (A) ORGANISME : singe

(ix) CARACTERISTIQUE ADDITIONNELLE :

- (A) NOM/CLE : CDS
- (B) EMPLACEMENT : 154..2064

(xi) DESCRIPTION DE LA SEQUENCE No 1 :

1 TGGCTGCCCCG CCGCGCGCACC CGCCCCGAGG CCTGTGCTCC TCGAAGGGG
 51 ACCGAGCGAA GCGGGGGCCC GCGCCAGGCC GCGCGGGACC GACCGCGATG
 101 CCGGGAGCTG CGACGGCTGC AGAGCGAGCT GCGCTCGGAG GCGGGTGTGA
 151 GGAAGATGGC CCAGTCCACC ACCACCTCCC CGGATGGGGG CACCACGTTT
 201 GAGCACCTCT GGAGCTCTCT GGAACCAGAC ACCACCTACT TCGACCTTCC
 251 CCAGTCAAGC CCGGGGAATA ATGAGGTGGT GGGTGGCAGC GATTCCAGCA
 301 TGGACGTCTT CCACCTAGAG GGCATGACCA CATCTGTCTT GCGCCAGTTC
 351 AATTTGCTGA GCAGCACCAT GGACCAGATG AGCAGCGCGC CTGCTCGGGC
 401 CAGCCCGTAC ACCCGGAGC ACGCCGCCAG CGTGGCCACC CATTACCCCT
 451 ACGCACAGCC CAGCTCCACC TTGACACCA TGTGCGCCGC GCCTGTCTTC
 501 CCCTCCAACA CCGACTATCC CGGACCCAC CACTTCGAGG TCACTTTCCA
 551 GCAGTCCAGC ACGGCCAAGT CAGCCACCTG GACGTACTCC CCACTCTTGA
 601 AGAAACTCTA CTGCCAGATC GCCAAGACAT GCGCCATCCA GATCAAGGTG
 651 TCCGCCCCAC CCGCCCCGGG CACCGCCATC CGGGCCATGC CTGTCTACAA
 701 GAAGGGCGGAG CACGTGACCG ACATCGTGAA GCGCTGCCCC AACCAAGAGC
 751 TCGGGAGGGA CTTCAACGAA GGACAGTCTG CCCCAGCCAG CCACCTCATC
 801 CGTGTGGAAG GCAATAATCT CTCGCAGTAT GTGGACGACC CTGTCAACGG
 851 CAGGCAGAGC GTCGTGGTGC CCTATGAGCC ACCACAGGTG GGGACAGAAT
 901 TCACCACCAT CCTGTACAAC TTCTGTGTA ACAGCAGCTG TGTGGGGGGC
 951 ATGAACCGAC GCGCCATCCT CATCATCATC ACCCTGAGCA CGCGGGATGG
 1001 GCAGGTGCTG GCGCGCCGGT CCTTCGAGGG CCGCATCTGC GCCTGTCTTG
 1051 GCGCGGACCG AAAAGCCGAT GAGGACCACT ACCGGGAGCA GCAGGCCTTG
 1101 AATGAGAGCT CCGCCAAGAA CGGGGCTGCC AGCAAGCGCG CCTTCAAGCA
 1151 GAGTCCCCCT GCGGTCCCCG CCCTGGGCCC GGGTGTGAAG AAGCGGCGGC
 1201 ACSGAGACGA GGACACCTAC TACCTGCAGG TCGAGGCGCG CGAGAAGTTC
 1251 GAGATCCTGA TGAAGCTGAA GGAGAGCCTG GAGCTGATGG AGTTGGTGCC
 1301 GCAGCCGCTG GTAGACTCCT ATCGGCAGCA GCAGCAGCTC CTACAGAGGC
 1351 CGAGTCACCT ACAGCCCCCA TCCTACGGGC CGGTCTCTTC GCGCATGAAC
 1401 AAGGTGCACG GGGGCGTGAA CAAGCTGCCC TCCGTCAACC AGCTGGTGGG
 1451 CCAGCCTCCC CCGCACAGCT CCGCAGCTAC ACCCAACCTG GGACCTGTGG
 1501 GCTCTGGGAT GCTCAACAAC CACGGCCACG CAGTGCCAGC CAACAGCGAG
 1551 ATGACCAGCA GCCACGGCAC CCAGTCCATG GTCTCGGGGT CCCACTGCAC
 1601 TCCGCCACCC CCTACCAGC CCGACCCAG CCTCGTCAGT TTTTAAACAG
 1651 GATTGGGGTG TCCAAACTGC ATCGAGTATT TCAGTCCCA GGGGTTACAG
 1701 AGCATTTACC ACCTGCAGAA CTTGACCATC GAGGACCTGG GGGCCCTGAA
 1751 GATCCCCGAG CAGTATCGCA TGACCATCTG GCGGGGCTG CAGGACCTGA
 1801 AGCAGGGCCA CGACTACGGC GCGCGCGCGC AGCAGCTGCT CCGCTCCAGC
 1851 AACCGGGCGG CCATTTCAT CCGCGGCTCC GGGGAGCTGC AGCGCCAGCG
 1901 GGTCTAGGAG GCGGTGCACT TCCGGGTGCG CCACACCATC ACCATCCCCA

1951 ACCGCGGCGG CCGCGGCGCC GCGCCCGACG AGTGGGCGGA CTTCGGCTTC
2001 GACCTGCCCC ACTGCAAGGC CCGCAAGCAG CCCATCAAGG AGGAGTTCAC
2051 GGAGGCGCGAG ATCCACTGAG GGGCCGGGCC CAGCCAGAGC CTGTGCCACC
2101 GCCCAGAGAC CCAGGCGCGC TCGCTCTCCT TCCTGTGTCC AAAACTGCCT
2151 CCGGAGGCGAG GGCTTCAGG CTGTGCGCGG GGAAAGGCAA GGTCCGGCCC
2201 ATGCCCCGGC ACCTCACCGC CCCCAGGAGA GGCCCAAGCA CCAAAGCCGC
2251 CTGCGGACAG CCTGASTCAC CTGCAGAACG TTCTGGAGCT GCCCTAATGC
2301 TGGGCTTGCG GGGCAGGGGC CGGCCCACTC TCAGCCCTGC CACTGCCGGG
2351 CGTGCTCCAT GGCAGGCGTG GGTGGGGACC GCAGTGTGAG CTCCGACCTC
2401 CAGGCCTCAT CCTAGAGACT CTGTCATCTG CCGATCAAGC AAGGTCCTTC
2451 CAGAGGAAAG AATCCTCTTC GCTGGTGGAC TGCCAAAAAG TATTTGCGA
2501 CATCTTTTGG TTCTGGAGAG TGGTGAGCAG CCAAGCGACT GTGTCTGAAA
2551 CACCGTGCAT TTTCAGGGAA TGTCCTAAC GGGCTGGGGA CTCTCTCTGC
2601 TGGACTTGGG AGTGGCCTTT GCGCCAGCA CACTGTATTC TCGGGGACCG
2651 CCTCCTTCCT GCGCCTAACA ACCACCAAAG TGTGTCTGAA ATTGGAGAAA
2701 ACTGGGGAAG GCGCAACCCC TCCCAGGTGC GGAAGGATC TGGTACCGCC
2751 TCGGCCAGTG CCCCTCAGCC TGGCCACAGT CACCTCTCCT TGGGGAACCC
2801 TGGGCAGAAA GGGACAGCCT GTCCTTAGAG GACTGGAAAT TGTCAATATT
2851 TGATAAAATG ATACCCCTTT CTAC

(2) INFORMATION POUR LA SEQ ID No 2 :

(i) CARACTERISTIQUES DE LA SEQUENCE :

- (A) LONGUEUR : 637 acides aminés
- (B) TYPE : acide aminé
- (D) CONFIGURATION : linéaire

(ii) TYPE DE MOLECULE : protéine

(xi) DESCRIPTION DE LA SEQ ID No 2 :

```

1  MAQSTTTSPD GGTTFEHLWS SLEPDSTYFD LPQSSRGNNV VGGTSSMD
51  VFHLEGMTTS VMAQFNLSS TMDQMSSRAA SASPYTPKHA ASVPTKSPYA
101 QPSSTFDTMS PAPVIPSNTD YPGPHHFEVT FQSSSTAKSA TWTYSPLLKK
151 LYCQIAKTCQ IQIKVSAPPP PGTAIRAMPV YKKAETHVDI VKRCRNHELQ
201 RDNFEGQSAP ASHLIRVEGN NLSQYVDDPV TGRQSVVVPY EPPQVGTEFT
251 TILYNFMCNS SCVGGMTRRP ILITITLLETR DGQVLGRRSF EGRICACPGR
301 DRKADEDHYR EQQALNESSA KNGAASKRAF KQSPPAVPAL GPGVYKRPHG
351 DEDTYYLQVR GRNFEILMK LKESLELMEL VPQPLVDSYR QQQQLLQRFS
401 HLQPPSYGPV LSPMKVHGG VKKLPSVNQL VGQPPPHSSA ATPNLGPVGS
451 GMLNNHGHAV PANSEMTSSH GTQSMVSGSH CTPPPPYHAD PSLVSFLTGL
501 GCPNCIEYFT SQGLQSIYHL QNLTIEDLGA LKIPPEQYRMT IWRGLQDLKQ
551 GHDYGAAAQQ LLRSSNAAAI SIGGSGELQR QRVMEAVHFR VFHTITIPNR
601 GGPAGAPDEN ADFGFDLPDC KARKQPIKEE FTEASIH

```

(2) INFORMATION POUR LA SEQ ID No 3 :

- (i) CARACTERISTIQUES DE LA SEQUENCE :
 - (A) LONGUEUR : 2034 paires de bases
 - (B) TYPE : acide nucléique
 - (C) NOMBRE DE BRINS : double
 - (D) CONFIGURATION : linéaire
- (ii) TYPE DE MOLECULE : ADNc
- (vi) ORIGINE :
 - (A) ORGANISME : singe
- (ix) CARACTERISTIQUE ADDITIONNELLE :
 - (A) NOM/CLE : CDS
 - (B) EMBLACEMENT : 154..1650
- (xi) DESCRIPTION DE LA SEQUENCE No 3 :

1 TGGCTGCTCCG CCGCGCGACC CGCGCGGAGG CCTGTGCTCC TCGAAGGGG
 51 ACCGAGCGAA GCGCGGCGCC GCGCGAGGCC GCGCGGACG GACGCGGATG
 101 CCGCGAGCTG CGACGGCTGC AGAGCGAGCT GCGCTCGGAG GCGGGTGTGA
 151 GGAAGATGGC CGAGTCCACT ACCACCTCCC CGGATGGGGG CACCACCTTT
 201 GAGCAGCTCT GGAGCTCTCT GGAACCGAGC AGCAGCTACT TCGACCTTCC
 251 CCAGTCAAGC CGGGGGAATA ATGAGGTGGT GGGTGGCAGC GATTCAGCA
 301 TGGACGTCTT CCACCTAGAG GGCATGACCA CATCTGTCTT GCGCCAGTTC
 351 AATTTGCTGA GCAGCACCAT GGACCAGATG AGCAGCGCGG CTGCGCTGGG
 401 CAGCGCTGAC ACCCGCGAGC ACGGCGCGAG CGTGGCCACC CATTACCGT
 451 ACCGACAGCG CAGCTCCACC TTCGACACCA TGTGCGCGCG GCGTGTCTC
 501 CCTCCAACA CCGACTATCC CGGACCGCAC CACTTCGAGG TCACCTTCCA
 551 GCAGTCCAGC ACGGCCAAGT CAGCCACCTG GACGTACTCC CCACTCTTGA
 601 AGAAACTCTA CTGCCAGATC GCGAAGACAT GCGCCATCCA GATCAAGGTG
 651 TCGCGCGCAC CGCGCGCGGG CACCGCGATC CGGGCGATGC CTGTCTACAA
 701 GAAGGCGGAG CAGCTGACCG ACATCTGAA GCGCTGCGCC AACCAGAGC
 751 TCGCGAGGGA CTTCACGAA GGACAGTCTG CCGCAGCGAG CACCTCATC
 801 CGTGTGGAAG GCAATAATCT CTGCGAGTAT GTGGACGACC CTGTACCGG
 851 CAGCGAGAGC GTCTGTGTC CCTATGAGCC ACCACAGGTG GGGACAGAAT
 901 TCACCACCAT CCGTACAAAC TTCATGTGTA ACAGCAGCTG TGTGGGGGGC
 951 ATGAACCGAC GCGCGATCTT CATCATCATC ACCCTGGAGA CGCGGGATGG
 1001 GCAGGTGCTG GCGCGCGGT CTTTCGAGGG CCGCATCTGC GCGTGTCTG
 1051 GCGCGGACCG AAAAGCGGAT GAGGACCACT ACCCGGAGCA GCAGGCGTTG
 1101 AATGAGAGCT CCGCGAAGAA CGGGGCTGCC AGCAAGCGCG CTTCAAGCA
 1151 GAGTCCCCCT GCGCTCCCCG CCGTGGCGCC GGGTGTGAAG AAGCGCGGGC
 1201 ACCGAGACGA GGACAGCTAC TACCTGCGAG TCGGAGGCGG CGAGAAGTTC
 1251 GAGATCCTGA TGAAGCTGAA GGAGAGCCTG GAGCTGATGG AGTTGGTGCC
 1301 GCAGCGCGCTG GTAGACTCTT ATCGGCAGCA GCAGCAGCTC CTACAGAGGC
 1351 CGAGTCACCT ACAGCGCGCA TCGTACGGCG CGGTCTCTTC GCGCATGAAC
 1401 AAGGTGCACG GGGGCGTGAA CAAGCTGCCC TCGTCAACC AGCTGGTGGG
 1451 CGAGCGTCCC CGGACAGCT CCGCAGCTAC ACCCAACCTG GGACCTGTGG
 1501 GCTCTGGGAT GCTCAACAAC CACGGCCACG CAGTGGCAGC CAACAGCGAG
 1551 ATGACCAGCA GCGACGGCAC CCAGTCCATG GTCTCGGGGT CCCACTGCAC
 1601 TCGCGCACCC CCTACCACG CCGACCGCAG CCGCTCAGG ACCTGGGGGC
 1651 CCGAAGATC CCGAGCAGT ATCGCATGAC CATCTGGCGG GCGCTGCAGG
 1701 ACCTGAAGCA GGGCGACGAC TACGGCGCGG CCGCGCAGCA CCGCTCCCGC
 1751 TCCAGCAACG CGGCGCGCAT TTCCATCGGC GCGTCCGGGG AGCTGCAGCG
 1801 CCAGCGGGTC ATGGAGCGCG TCGACTTCCG CGTGGCGCAC ACCATCACCA
 1851 TCCCAACCG CGGCGCGCGG GCGCGCGCG CCGACGAGTG GCGGAGTTC
 1901 GCGTTCGACC TCGCGGACTG CAAGGCGCGC AAGCAGCTCA TCAAGGAGGA

1951 GTTCACGGAG GCCGAGATCC ACTGAGGGGC CGGGCCCAGC CAGAGCCTGT

2001 GCCACCGCCC AGAGACCCAG GCCGCCTCGC TCTC

(2) INFORMATION POUR LA SEQ ID No 4 :

(i) CARACTERISTIQUES DE LA SEQUENCE :

- (A) LONGUEUR : 499 acides aminés
- (B) TYPE : acide aminé
- (D) CONFIGURATION : linéaire

(ii) TYPE DE MOLECULE : protéine

(xi) DESCRIPTION DE LA SEQ ID No 4 :

```

1  MAQSTTTSPD GDTTFEHLWS SLEPDSTYFD LQSSRCNNE VVGCTDSSMD
51  VFKLEGMTTS VMAQFNLLSS TMEQMSSRAA SASPTTPEHA ASVPTKSPYA
101 QPSSTFDTMS PAPVIPSNTD YPGPHNFVET FQSSSTAKSA TWTYSPLLKK
151 LYCQIAKTCF IQIKVSAPPP PGTAIRAMPV YKQAKVTDI VKRCPNHELG
201 RDRNEGQSAP ASHLIRVEGN NLSQIVDDPV TGRQSVVVPY EPPQVGTEFT
251 TILYNFYONS SCVGGYRPP ILIIITLETR DGQVLGRASF EGRICACPR
301 DRKADECHYR EQQALNESSA KNGAASKRAF KQSPPAVPAL GPGVKKRRHG
351 DEDTY/LQVR GRENFELMK LKESLELMEL VPQPLVDSYR CQCQLLQRPS
401 HLQPPSYQPV LSPYKX/HOG VNKLPVNIQL VQQPPPHSSA ATPNLGPVGS
451 GMLNTHGHAV PANSEMTSSH GTQSMVSGSH CT????PYHAD PSLVRTWGP

```

- (2) INFORMATION POUR LA SEQ ID No 5 :
- (i) CARACTERISTIQUES DE LA SEQUENCE :
 - (A) LONGUEUR : 2156 paires de bases
 - (B) TYPE : acide nucléique
 - (C) NOMBRE DE BRINS : double
 - (D) CONFIGURATION : linéaire
 - (ii) TYPE DE MOLECULE : ADNC
 - (vi) ORIGINE :
 - (A) ORGANISME : homme
 - (ix) CARACTERISTIQUE ADDITIONNELLE :
 - (A) NOM/CLE : CDS
 - (B) EMPLACEMENT : 33..1940
 - (xi) DESCRIPTION DE LA SEQUENCE No 5 :

1 GCGAGCTGCC CTCGAGGCC GGCCTGGGA AGATGGGCA GTCCACCGCC
 51 ACCTCCCTCTG ATGGGGGCAC CACGTTTGAG CACCTCTGGA GCTCTCTGGA
 101 ACCAGACAGC ACCTACTTCC ACCTTCCCA GTCAAGCCCG GGGAAATAATG
 151 AGGTGGTGGG GGGAACTGAT TCCAGCATGG AGCTCTTCCA CCTGGAGGGC
 201 ATGACTACAT CTGTCTATGG CCAGTTCAAT CTGCTGAGCA GCACCATGGA
 251 CCAGATGAGC AGCCGCGCGG CCTCGGCCAG CCCCTACACC CCAGAGCAGC
 301 CCGCCAGCGT GCCACCCAC TCGCCCTACG CACAACCCAG CTCCACCTTC
 351 GACACCATGT CGCCGGCGCC TGTCATCCCC TCCAACACCG ACTACCCCGG
 401 ACCCCACCAC TTTGAGGTCA CTTTCTAGCA GTCCAGCAGC GCCAAGTCAG
 451 CCACCTGGAC GTACTCCCGC CTCTTGAAGA AACTCTACTG CCAGATCGCC
 501 AAGACATGCC CCATCCAGAT CAAGGTGTCC ACCCCGCCAC CCCCAGGCAC
 551 TGGCATCCCG GCCATGCGTG TTTACAAGAA AGCGGAgcAC GTGACCGAGC
 601 TCGTGAACCG CTGCCCCAAC CACGAGCTCG GGAGGGACTT CAACGAAGGA
 651 GATCTGTCTC GAGCCAGCCA CCTCATCCGC GTGGAAAGCA ATAATCTCTC
 701 GCAGTATGTG GATGACCTTG TCACCGGCAG GCAGAGCGTC GTGGTGCCTT
 751 ATGAGCCACC ACAGGTGGGG ACGGAAATCA CCACCATGCT GTACAACTTC
 801 ATGTGTAACA GCAGCTGTGT AGGGGGCATG AACGGCGCGC CCATCCTCAT
 851 CATCATCACC CTGGAGATGC GGGATGGGCA GGTGCTGGGC CGCCGGTCTT
 901 TTGAGGGCCG CATCTGCGCC TGTCTTGGCC GCGACCGAAA AGCTGATGAG
 951 GACCACTACC GGGAGCAGCA GGCCTGAAAC GAGAGCTCCG CCAAGAACCG
 1001 GGCCGCCAGC AAGCTGCGT TCAAGCAGAG CCCCCGTGCC GTCCCGCGCC
 1051 TTGGTGGCGG TGTGAAGAAg CGCGCGCATG GAGACGAGGA CACGTACTAC
 1101 CTTACAGTGC GAGGCCGGGA GAACTTTGAG ATCTTGATGA AGCTGAAAGA
 1151 GAGCTGGAG CTGATGAGT TGGTGGGCA GCGACTGGTG GACTCTATC
 1201 GGCAGCAGCA GCAGCTCTTA CAGAGGCGGA GTCACTACA GCGCCCGTCC
 1251 TACGGGCGCG TCTCTCGCC CATGAACAAG GTGCACGGGG GCATGAACAA
 1301 CCTGCCCTCC GTCAACCAGC TGGTGGGCA GCTTCCCCCG CACAGTTCCG
 1351 CAGCTACACC CAACCTGGGG CCGCTGGGCC CCGGGATGCT CAACAACCAT
 1401 GGCCACGCGG TGCCAGCCAA CGCGGAGATG AGCAGCAGCC ACAGCGCCCA
 1451 GTCCATGGTC TCGGGGTCCC ACTGCACTCC GCCACCGCCC TACCAGCGCG
 1501 ACCCCAGCGT CGTCACTTTT TTAACAGGAT TGGGGTGTCC AAACCTGCATC
 1551 GAGTATTTCA CTTCCCAAGG GTTACAGAGC ATTTACCACC TgcAGAACcT
 1601 GACcATTGAG GACCTGGGGG CCTGGAAGAT CCCCAGGCAG TACCGCATGA
 1651 CCATCTGGCG GGGCTGCGAG GACCTGAAGC AGGGCCACGA CTACAGCACC
 1701 GCGCAGCAGC TGCTCCGCTC TAGCAACGCG GCCACCATCT CCATCGGCGG
 1751 CTCAGGGGAA CTGCAGCGCC AGCGGGTCAT GGAGGCGGTG CACTTCCCGG
 1801 TGCGCCACAC CATCACCATC CCAAACCGCG GCGGCGCAGG CGGCGGCCCT
 1851 GACGAGTGGG CGGACTTCGG CTTCGACCTG CCGGACTGCA AGGCCCGCAA
 1901 GCAGCCCATC AAGGAGGAGT TCACGGAGGC CGAGATCCAC TGAGGGCCTC

1951 GCCTGGCTGC AGCCTGCGCC ACCGCCCAGA GACCCAAGCT GCCTCCCCCTC
2001 TCCTTCCTGT GTGTCCAAAA CTGCCTCAGG AGGCAGGACC TTCGGGCTGT
2051 GCGCGGGGAA AGGCAAGGTC CGGCCCCATCC CCAGGCACCT CACAGGCCCC
2101 AGGAAAGGCC CAGCCACCGA AGCCGCCTGT GGACAGCCTG AGTCACCTGC
2151 AGAACC

(2) INFORMATION POUR LA SEQ ID No 6 :

(i) CARACTERISTIQUES DE LA SEQUENCE :

- (A) LONGUEUR : 636 acides aminés
- (B) TYPE : acide aminé
- (D) CONFIGURATION : linéaire

(ii) TYPE DE MOLECULE : protéine

(xi) DESCRIPTION DE LA SEQ ID No 6 :

```

1  MAQSTATSPD GGTTFEHLWS SLEPDSTYFD LPQSSRGNNF VGGTSSMD
51  VFHLEGMTTS VMAQFNLLSS TMDQMSSRAA SASPYTPEHA ASVPTHSPYA
101 QPSSTFDCTS PAPVIPSNTD YPGPHHFEVT FQSSSTAKSA TWTYSPLLKK
151 LYCQIAKTCP IQIKVSTPPP PGTAIRAMPV YKKAEMVTDV VKRCPIHELG
201 EDFNEGQSAF ASHLIRVEGN NLSQVVDOPV TGRQSVVVPY EPPQVGTEFT
251 TILVNFMCNS SCVGGMNRRP ILIIITLENR DGQVLGRSF EGRICACPR
301 DRKADEDHYR EQQALNESSA KNGAASKRAF KQSPPAVPAL GAGVKKRRHG
351 DEDTYYLQVR GRENFELMK LKESLELMEL VPQPLVDSYR QCCQLLQRPS
401 HLQPPSYGPV LSPMKVHGG MKKLPSVNQL VQPPPHSSA ATPNLGPVGP
451 GMLNNHGHAV PANGEMSSSH SAQSMVSGSH CTPPPPYHAD PSLVSFLTGL
501 GCPNCIEYFT SQGLQSIYHL QNLTIEDLGA LKIQEQYRMT IWRGLQDLKQ
551 GRDYSTAQQL LRSSNAATIS IGGSGELQRQ RVMEAVHFRV PHTITIPNRG
601 GPGGGPDEWA DFGFDLPDCK ARKQPIKEEF TEASIH

```

(2) INFORMATION POUR LA SEQ ID No 7 :

(i) CARACTERISTIQUES DE LA SEQUENCE :

- (A) LONGUEUR : 2040 paires de bases
- (B) TYPE : acide nucléique
- (C) NOMBRE DE BRINS : double
- (D) CONFIGURATION : linéaire

(ii) TYPE DE MOLECULE : ADNc

(vi) ORIGINE :

- (A) ORGANISME : souris

(ix) CARACTERISTIQUE ADDITIONNELLE :

- (A) NOM/CLE : CDS
- (B) EMPLACEMENT : 124..1890

(xi) DESCRIPTION DE LA SEQUENCE No 7 :

1 TGAATCTCCCT GTGGGCTGCA GGGGACTGAG CCAGGGAGTA GATGCCCTGA
 51 GAGCCCAAGG GACACCAAG GAAACCTTGC TGGCTTTGAG AAAGGGATCG
 101 TGTCTCTCCCT GGGCAAGAA AGCATGTGTA TGGGCCCCGT GTATGAATCC
 151 TTGGGGCAGG CCCAGTTCAA TTTGCTCAGC AGTGGCATGG ACCAGATGGG
 201 CAGCCGTGCG GCGCGGCGA GCGGCTACAC CCGGAGGAC GCGGCCAGCG
 251 CGCCCAACCA CTCGCGCTAC GCGCAGCCCA GCTCCACCTT CGACACCATG
 301 TCTCCGGGGC CTGTCTCCG TTCCAATACC GACTACCCCG GCGCCACCA
 351 CTTCCAGGTC ACCTTCCAGC AGTCGAGCAC TGCCAAGTCG GCCACCTGGA
 401 CATACTCCCC ACTCTTGAAG AAGTTGTACT GTCAGATTGC TAAGACATGC
 451 CCCATCCAGA TCAAGTGTG CACACCACCA CCCCCGGCA CGGCCATCCG
 501 GGGCATGCCCT GTCTACAAGA AGGCAGAGCA TGTGACCGAC ATTGTTAAGC
 551 GCTGCCCCAA CCACGAGCTT GGAAGGGACT TCAATGAAGG ACAGTCTGCC
 601 CCGGCTAGCC ACCTCATCCG TGTAGAAGGC AACAACCTCG CCCAGTACGT
 651 GGATGACCTT GTCACCGAA GGCAGAGTGT GGTGTGCGG TATGAACCCC
 701 CACAGGTGGG AACAGATTT ACCACCATCC TGTACAACCT CATGTGTAAC
 751 AGCAGCTGTG TGGGGGGCAT GAATCGGAGG CCCATCCTTG TCATCATCAC
 801 CCTGGAGACC CCGGATGGAC AGGTCTGGG CCGCGGTCTT TTCGAGGGTC
 851 GCATCTGTGC CTGTCTGGG CTGACCGCA AAGCTGATGA AGACCATTAC
 901 CCGGAGCAAC AGGCTCTGAA TGAAGTACC ACCAAAAATG GAGCTGCCAG
 951 CAAAGCTGCA TTCAAGCAGA GCGGCGCTGC CATCCCTGCC CTGGGTACCA
 1001 ACSTGAAGAA GAGACCCAC GGGGACGAGG ACATGTTCTA CATGCACGTG
 1051 CGAGGCCGGG AGAAGTTTGA GATCTTGATG AAAGTCAAGG AGAGCCTAGA
 1101 ACTGATGGAG CTTGTGCCCC AGCCTTTGGT TGAATCCTAT CGACAGCAGC
 1151 AGGAGCAGCA GCTCTACAG AGGCGGAGTC ACCTGCAGCC TCCATCCTAT
 1201 GGGGCGCTGC TCTCCCAAT GAACAAGGA CAGGCTGGTG TCAACAACT
 1251 GCGCTCTCTG AACCAAGCTG TGGGCCAGCC TCGCCCGCAC AGCTCAGCAG
 1301 CTGGGCCCCA CTTGGGGCCC ATGGGCTCCG GGATGCTCAA CAGCCACGGC
 1351 CACAGCATGC CGGCAATGG TGAGATGAAT GGAGGCCACA GCTCCAGAC
 1401 CATGGTTTGG GGATCCCACT GACCCCGCC ACCCCCTAT CATGCAGACC
 1451 CCAGCCTCTT CAGTTTTTTG ACAGGGTTGG GGTGTCCAAA CTGCATCGAG
 1501 TGCTTCACTT CCAAAGGGTT GCAGAGCATC TACCACCTGC AGAACCTTAC
 1551 CATCGAGGAC CTTGGGGGTC TGAAGGTCCC TGACCAGTAC CGTATGACCA
 1601 TCTGGAGGGG CTTACAGGAC CTGAAGCAGA GCCATGACTG CGGCCAGCAA
 1651 CTGCTACGCT CCAGCAGCAA CGCGGCCACC ATCTCCATCG GCGGCTCTGG
 1701 CGAGCTGCAG CCGCAGCGGG TCATGGAAGC CGTGCAATTC CGTGTGGGCC
 1751 ACACCATCAC AATCCCCAAC CGTGGAGGCG CAGGTGCGGT GACAGGTCCC
 1801 GACGAGTGGG CGGACTTTGG CTTTGACCTG CTTGACTGCA AGTCCCGTAA
 1851 GCAGCCCATC AAAGAGGAGT TCACAGAGAC AGAGAGCCAC TGAGGAACGT
 1901 ACCTTCTCTT CTTGTCTTTC CTCTGTGAGA AACTGCTCTT GGAAGTGGGA

1951 CCTGTTGGCT GTGCCCACAG AAACCAGCAA GGACCTTCTG CCGATGCCA
2001 TTCTGAAGG GAACTCGCTC ATGAACCTAC TCCCTCTTGG

(2) INFORMATION POUR LA SEQ ID No 8 :

(i) CARACTERISTIQUES DE LA SEQUENCE :

- (A) LONGUEUR : 589 acides aminés
- (B) TYPE : acide aminé
- (D) CONFIGURATION : linéaire

(ii) TYPE DE MOLECULE : protéine

(xi) DESCRIPTION DE LA SEQ ID No 8 :

```

1  MCMGPVYESL GQAQFNLLSS AMQMGSRRA PASPYTPERA ASAPTHSPYA
51  QPSSTFDCHS PAPVIPSNVD YPGPHFEVT FQSSSTAKSA TWTYSPLLKK
101 LYCQIAKTOP IQIKVSTPPP PGTAIPAMPV YKKAHWTDI VKRCPNHELQ
151 PDFNEGQSAP ASHLIRVEGN NLAQVDDPV TGRQSVVVPY EPPQVGTEFT
201 TILVNFMCNS SCVGGMRBP ILVITLLETR DGQVLGRSF EGRICACPR
251 DRKADECHYR EQQALNESTT KNGAASKRAF KQSPPAIPAL GTNVKKRRHG
301 DEDMFYHVR GRNFELMK VKESLELMEL VPQPLVDSYR QQQQQQLLQR
351 PSHLQPPSYG PVLSPYKXVH GGVNKLPSVN QLVGQPPPHS SAAGPNLQPM
401 GSQMLNSGHG SMPANGEMNG GHSSQTMVSG SHCTPPPPYH ADPSLVSLT
451 GLGCPNCIEC FTSQGLQSTY HLQMLTIEDL GALKVPDQYR MTINRGLQDL
501 KQSHDCGQQL LRSSNAATI SIGSGELQR QRMEAVHFR VRLTITIPNR
551 GGAGAVTOPD ENADFGFDLP DCKSRKQPIK EFTTETESH

```

(2) INFORMATION POUR LA SEQ ID No 9 :

(i) CARACTERISTIQUES DE LA SEQUENCE :

- (A) LONGUEUR : 758 paires de bases
- (B) TYPE : acide nucléique
- (C) NOMBRE DE BRINS : double
- (D) CONFIGURATION : linéaire

(ii) TYPE DE MOLECULE : ADNC

(vi) ORIGINE :

(A) ORGANISME : souris

(ix) CARACTERISTIQUE ADDITIONNELLE :

- (A) NOM/CLE : CDS
- (B) EMBLEMMENT : 388..756

(xi) DESCRIPTION DE LA SEQUENCE No 9 :

```

1  TGGTCCCGCT TCGACCAAGA CTCCGGGTAC CAGCTTGCAG GCCCCGCGGA
51  GGAGGAGACC CCGCTGGGGC TAGCTGGGCG ACGCGCGCCA AGCGGCGGCG
101 GGAAGGAGGC GGGAGGAGCG GGGCCCGAGA CCCGACTCG GGCAGAGCCA
151 GCTGGGGAGG CCGGGCCCGC GTGGGAGCCA GGGGCCCGGG TGGCCGGCCC
201 TCCTCCGCCA CCGCTGAGTG CCGCGGCTGC CTCCCGCGCG GTCCGCCAAG
251 AAAGGCGCTA AGCCTGCGGC AGTCCCGCTG CCGCGCGCTC CCTGCTCCGC
301 ACCCTTATAA CCGCCCGTCC CGCATCCAGG CGAGGAGGCA ACGCTGCAGC
351 CCAGCCCTCG CCGACGCGGA CGCCCGGCCC GGAGCAGAAT GAGCGGCAGC
401 GTTGGGGAGA TGGCCGAGAC CTCTTCTTCC TCCTCCTCCA CCTTCGAGCA
451 CCGTGTGGAGT TCTCTAGAGC CAGACAGCAC CTACTTTGAC CTCGCCCAGC
501 CCGAGCCAAGG GACTAGCGAG GCATCAGGCA GCGAGGAGTC CAACATGGAT
551 GTCTTCCACC TGCAAGGCAT GGGCCAGTTC AATTTGCTCA GCAGTGCCAT
601 GGATCAGATG GGCAGCCGTG CCGCCCGGCG GAGCCCGTAC ACCCCGGAGC
651 ACGCCGCGAG CCGCGCCGAC CACTCGGCGT ACSCGCAGCC CAGCTCCACC
701 TTCGACACCA TGTCTCGGC GCCTGTCATC CCTTCCAATA CCGACTACCC
751 CGGCCGCC

```

- (2) INFORMATION POUR LA SEQ ID No 10 :
- (i) CARACTERISTIQUES DE LA SEQUENCE :
 - (A) LONGUEUR : 123 acides aminés
 - (B) TYPE : acide aminé
 - (D) CONFIGURATION : linéaire
 - (ii) TYPE DE MOLECULE : protéine
 - (xi) DESCRIPTION DE LA SEQ ID No 10 :

1 MSGSVGEMAQ TSSSSSSSTFE RLWSSLEPDS TYFDLPQPSQ GTSEASGSEE
51 SNMDEVTHLQG MAQFNLLSSA MDQMGSRAAP ASPYTPEHAA SAPTHSPYAQ
101 PSSTFDCTMS? APVIPENTDY PG?

- (2) INFORMATION POUR LA SEQ ID No 11 :
- (i) CARACTERISTIQUES DE LA SEQUENCE :
 - (A) LONGUEUR : 17 paires de bases
 - (B) TYPE : acide nucléique
 - (C) NOMBRE DE BRINS : simple
 - (D) CONFIGURATION : linéaire
 - (ii) TYPE DE MOLECULE : oligonucléotide
 - (iii) ANTI-SENS : non
 - (xi) DESCRIPTION DE LA SEQ ID No 11 :

GCG AGC TGC CCT CGG **AG**

(2) INFORMATION POUR LA SEQ ID No 12 :

(i) CARACTERISTIQUES DE LA SEQUENCE :

- (A) LONGUEUR : 19 paires de bases
- (B) TYPE : acide nucléique
- (C) NOMBRE DE BRINS : simple
- (D) CONFIGURATION : linéaire

(ii) TYPE DE MOLECULE : oligonucléotide

(iii) ANTI-SENS : oui

(xi) DESCRIPTION DE LA SEQ ID No 12 :

GGT TCT GCA GGT GAC **TCA G**

- (2) INFORMATION POUR LA SEQ ID No 13 :
- (i) CARACTERISTIQUES DE LA SEQUENCE :
 - (A) LONGUEUR : 18 paires de bases
 - (B) TYPE : acide nucléique
 - (C) NOMBRE DE BRINS : simple
 - (D) CONFIGURATION : linéaire
 - (ii) TYPE DE MOLECULE : oligonucléotide
 - (iii) ANTI-SENS : non
 - (xi) DESCRIPTION DE LA SEQ ID No 13 :

GCC ATG CCT GTC TAC AAG

(2) INFORMATION POUR LA SEQ ID No 14 :

(i) CARACTERISTIQUES DE LA SEQUENCE :

- (A) LONGUEUR : 18 paires de bases
- (B) TYPE : acide nucléique
- (C) NOMBRE DE BRINS : simple
- (D) CONFIGURATION : linéaire

(ii) TYPE DE MOLECULE : oligonucléotide

(iii) ANTI-SENS : oui

(xi) DESCRIPTION DE LA SEQ ID No 14 :

ACC AGC TGG TTG ACG GAG

(2) INFORMATION POUR LA SEQ ID No 15 :

(i) CARACTERISTIQUES DE LA SEQUENCE :

- (A) LONGUEUR : 21 paires de bases
- (B) TYPE : acide nucléique
- (C) NOMBRE DE BRINS : simple
- (D) CONFIGURATION : linéaire

(ii) TYPE DE MOLECULE : oligonucléotide

(iii) ANTI- SENS: non

(xi) DESCRIPTION DE LA SEQ ID No 15 :

GTC AAC CAG CTG GTG GGC CAG

- (2) INFORMATION POUR LA SEQ ID No 16 :
- (i) CARACTERISTIQUES DE LA SEQUENCE :
 - (A) LONGUEUR : 16 paires de bases
 - (B) TYPE : acide nucléique
 - (C) NOMBRE DE BRINS : simple
 - (D) CONFIGURATION : linéaire
 - (ii) TYPE DE MOLECULE : oligonucléotide
 - (iii) ANTI-SENS : oui
 - (xi) DESCRIPTION DE LA SEQ ID No 16 :
GTG GAT CTC GGC CTC C

REVENDICATIONS

1. Polypeptide purifié, comprenant une séquence d'acides aminés choisie parmi :
 - a) la séquence SEQ ID n° 2 ;
 - b) la séquence SEQ ID n° 4 ;
 - c) la séquence SEQ ID n° 6 ;
 - d) la séquence SEQ ID n° 8 ;
 - e) la séquence SEQ ID n° 10 ;
 - f) toute séquence biologiquement active dérivée de SEQ ID n° 2, SEQ ID n° 4, SEQ ID n° 6, SEQ ID n° 8 ou SEQ ID n° 10.
2. Polypeptide selon la revendication 1, caractérisé en ce qu'il comprend la séquence d'acides aminés SEQ ID n° 6.
3. Polypeptide selon la revendication 1, caractérisé en ce qu'il comprend la séquence comprise entre :
 - le résidu 110 et le résidu 310 de SEQ ID n° 2, 4 ou 6 ;
 - le résidu 60 et le résidu 260 de SEQ ID n° 8 ;
 - le résidu 109 et le résidu 123 de SEQ ID n° 10.
4. Polypeptide selon la revendication 1, caractérisé en ce qu'il résulte d'un épissage alternatif de l'ARN messager du gène correspondant.
5. Polypeptide selon l'une quelconque des revendications précédentes, caractérisé en ce qu'il s'agit d'un polypeptide recombinant produit sous la forme d'une protéine de fusion.
6. Séquence d'acides nucléiques isolée codant pour un polypeptide selon l'une quelconque des revendications précédentes.
7. Séquence d'acides nucléiques isolée selon la revendication 6, caractérisée en ce qu'elle est choisie parmi :
 - a) la séquence SEQ ID n° 1 ;
 - b) la séquence SEQ ID n° 3 ;
 - c) la séquence SEQ ID n° 5 ;

- d) la séquence SEQ ID n°7 ;
 - e) la séquence SEQ ID n°9 ;
 - f) les séquences d'acides nucléiques capables de s'hybrider à la séquence SEQ ID n° 1, SEQ ID n° 3, SEQ ID n° 5, SEQ ID n°7 ou SEQ ID n°9 ou à leurs séquences complémentaires, ou de s'hybrider à leurs séquences proximales.;
 - g) les séquences dérivées des séquences a), b), c), d), e) ou f) du fait de la dégénérescence du code génétique.
- 5
- 10 8. Séquence nucléotidique selon la revendication 6, caractérisée en ce qu'il s'agit de la séquence SEQ ID n° 5 codant pour le polypeptide de séquences SEQ ID n° 6.
- 15 9. Vecteur de clonage et/ou d'expression contenant une séquence d'acides nucléiques selon l'une quelconque des revendications 6 à 8.
10. Vecteur selon la revendication 9, caractérisé en ce qu'il s'agit du plasmide pSE-1.
- 20 11. Cellule hôte transfectée par un vecteur selon la revendication 9 ou 10.
12. Cellule hôte transfectée selon la revendication 11, caractérisée en ce qu'il s'agit de *E. coli* MC 1061.
- 25 13. Sonde nucléotidique ou amorce nucléique caractérisée en ce qu'elle s'hybride spécifiquement avec l'une quelconque des séquences selon les revendications 6 à 8 ou leurs séquences complémentaires ou les ARN messagers correspondants ou les gènes correspondants.
- 30 14. Sonde selon la revendication 13, caractérisée en ce qu'elle comporte au moins 10 nucléotides.
- 35 15. Sonde selon la revendication 13, caractérisée en ce qu'elle comprend l'intégralité de la séquence du gène codant pour l'un des polypeptides de la revendication 1.

16. Sondes nucléotides caractérisées en ce qu'elles comprennent les oligonucléotides suivants ou leurs complémentaires :

SEQ ID n° 11 : GCG AGC TGC CCT CGG AG
 SEQ ID n° 12 : GGT TCT GCA GGT GAC TCA G
 SEQ ID n° 13 : GCC ATG CCT GTC TAC AAG
 SEQ ID n° 14 : ACC AGC TGG TTG ACG GAG
 SEQ ID n° 15 : GTC AAC CAG CTG GTG GGC CAG
 SEQ ID n° 16 : GTG GAT CTC GGC CTC C

17. Couples d'amorces nucléotidiques caractérisées en ce qu'elles s'hybrident avec l'un des séquences 6 à 8 ou leur complémentaire et permettent l'amplification d'une des séquences 6 à 8 par la technique de PCR ou toute autre variante de celle-ci.

18. Couple d'amorces nucléotidiques, caractérisé en ce qu'il comprend les amorces de séquences suivantes :

a) SEQ ID n° 11
 amorce sens : GCG AGC TGC CCT CGG AG

SEQ ID n° 12
 amorce antisens : GGT TCT GCA GGT GAC TCA G

-b) SEQ ID n° 13
 amorce sens : GCC ATG CCT GTC TAC AAG

SEQ ID n° 14
 amorce antisens : ACC AGC TGG TTG ACG GAG

c) SEQ ID n° 15
 amorce sens : GTC AAC CAG CTG GTG GGC CAG

SEQ ID n° 16
 amorce antisens : GTG GAT CTC GGC CTC C

19. Utilisation d'une séquence selon l'une quelconque des revendications 6 à 8, pour la réalisation de sondes nucléotidiques de diagnostic ou de séquences antisens utilisables en thérapie génique.
- 5 20. Utilisation d'une sonde selon l'une quelconque des revendications 13 à 16, comme outil de diagnostic *in vitro* pour la détection, par des expériences d'hybridation, des séquences d'acides nucléiques codant pour un polypeptide selon l'une quelconque des revendications 1 à 4, dans des échantillons biologiques, ou pour la mise en évidence de synthèses aberrantes ou d'anomalies génétiques.
- 10 21. Méthode de diagnostic *in vitro* pour la détection de synthèses aberrantes ou d'anomalies génétiques au niveau des séquences d'acides nucléiques codant pour un polypeptide selon l'une quelconque des revendications 1 à 4, caractérisée en ce qu'elle comprend :
- 15 – la mise en contact d'une sonde nucléotidique selon l'une quelconque des revendications 13 à 16 avec un échantillon biologique dans des conditions permettant la formation d'un complexe d'hybridation entre ladite sonde et la susdite séquence nucléotidique, éventuellement après une étape préalable d'amplification de la susdite séquence nucléotidique ;
- 20 – la détection du complexe d'hybridation éventuellement formé ;
- éventuellement le séquençage de la séquence nucléotidique formant le complexe d'hybridation avec la sonde de l'invention.
- 25 22. Utilisation d'une séquence d'acides nucléiques selon l'une quelconque des revendications 6 à 8, pour la production d'un polypeptide recombinant selon l'une quelconque des revendications 1 à 5.
- 30 23. Méthode de production d'une protéine recombinante SR-p70, caractérisée en ce que l'on cultive des cellules transfectées selon la revendication 10 ou 11 dans des conditions permettant l'expression d'un polypeptide recombinant de séquence SEQ ID n° 2, SEQ ID n° 4, SEQ ID n° 6, SEQ ID n° 8 ou SEQ ID n° 10 ou tout fragment ou dérivé biologiquement actif, et que l'on récupère ledit polypeptide recombinant.
- 35

16. Sondes nucléotides caractérisées en ce qu'elles comprennent les oligonucléotides suivants ou leurs complémentaires :
- SEQ ID n° 11 : GCG AGC TGC CCT CGG AG
 - SEQ ID n° 12 : GGT TCT GCA GGT GAC TCA G
 - SEQ ID n° 13 : GCC ATG CCT GTC TAC AAG
 - SEQ ID n° 14 : ACC AGC TGG TTG ACG GAG
 - SEQ ID n° 15 : GTC AAC CAG CTG GTG GGC CAG
 - SEQ ID n° 16 : GTG GAT CTC GGC CTC C
17. Utilisation d'une séquence **selon l'une** quelconque des revendications 6 à 8 pour la fabrication d'amorces oligonucléotidiques pour des réactions de séquençage ou d'amplification **spécifique** selon la technique de PCR ou toute variante de celle-ci.
18. Couple d'amorces nucléotidiques, caractérisé en ce qu'il comprend les amorces de séquences suivantes :
- a) SEQ ID n° 11
amorce sens : GCG AGC TGC CCT CGG AG
 - SEQ ID n° 12
amorce antisens : GGT TCT GCA GGT GAC TCA G
 - b) SEQ ID n° 13
amorce sens : GCC ATG CCT GTC TAC AAG
 - SEQ ID n° 14
amorce antisens : ACC AGC TGG TTG ACG GAG
 - c) SEQ ID n° 15
amorce sens : GTC AAC CAG CTG GTG GGC CAG
 - SEQ ID n° 16
amorce antisens : GTG GAT CTC GGC CTC C

19. Utilisation d'une séquence selon l'une quelconque des revendications 6 à 8, pour la réalisation de sondes nucléotidiques de diagnostic ou de séquences antisens utilisables en thérapie génique.
20. Utilisation d'une sonde selon l'une quelconque des revendications 13 à 16, comme outil de diagnostic *in vitro* pour la détection, par des expériences d'hybridation, des séquences d'acides nucléiques codant pour un polypeptide selon l'une quelconque des revendications 1 à 4, dans des échantillons biologiques, ou pour la mise en évidence de synthèses aberrantes ou d'anomalies génétiques.
21. Méthode de diagnostic *in vitro* pour la détection de synthèses aberrantes ou d'anomalies génétiques au niveau des séquences d'acides nucléiques codant pour un polypeptide selon l'une quelconque des revendications 1 à 4, caractérisée en ce qu'elle comprend :
 - la mise en contact d'une sonde nucléotidique selon l'une quelconque des revendications 13 à 16 avec un échantillon biologique dans des conditions permettant la formation d'un complexe d'hybridation entre ladite sonde et la susdite séquence nucléotidique, éventuellement après une étape préalable d'amplification de la susdite séquence nucléotidique ;
 - la détection du complexe d'hybridation éventuellement formé ;
 - éventuellement le séquençage de la séquence nucléotidique formant le complexe d'hybridation avec la sonde de l'invention.
22. Utilisation d'une séquence d'acides nucléiques selon l'une quelconque des revendications 6 à 8, pour la production d'un polypeptide recombinant selon l'une quelconque des revendications 1 à 5.
23. Méthode de production d'un polypeptide selon l'une quelconque des revendications 1 à 5 recombinant, caractérisée en ce que l'on cultive des cellules transfectées selon la revendication 11 ou 12 dans des conditions permettant l'expression d'un polypeptide recombinant de séquence SEQ ID n° 2, SEQ ID n° 4, SEQ ID n° 6, SEQ ID n° 8 ou SEQ ID n° 10 ou tout fragment ou dérivé biologiquement actif, et que l'on récupère ledit polypeptide recombinant.

16. Sondes nucléotides ayant les séquences suivantes ou leurs complémentaires :

SEQ ID n° 11 : GCG AGC TGC CCT CGG AG

SEQ ID n° 12 : GGT TCT GCA GGT GAC TCA G

SEQ ID n° 13 : GCC ATG CCT GTC TAC AAG

SEQ ID n° 14 : ACC AGC TGG TTG ACG GAG

SEQ ID n° 15 : GTC AAC CAG CTG GTG GGC CAG

SEQ ID n° 16 : GTG GAT CTC GGC CTC C

17. Utilisation d'une séquence selon l'une quelconque des revendications 6 à 8 pour la fabrication d'amorces oligonucléotidiques pour des réactions de séquençage ou d'amplification spécifique selon la technique de PCR ou toute variante de celle-ci.

18. Couple d'amorces nucléotidiques, caractérisé en ce qu'il comprend les amorces de séquences suivantes :

a) SEQ ID n° 11

amorce sens : GCG AGC TGC CCT CGG AG

SEQ ID n° 12

amorce antisens : GGT TCT GCA GGT GAC TCA G

b) SEQ ID n° 13

amorce sens : GCC ATG CCT GTC TAC AAG

SEQ ID n° 14

amorce antisens : ACC AGC TGG TTG ACG GAG

c) SEQ ID n° 15

amorce sens : GTC AAC CAG CTG GTG GGC CAG

SEQ ID n° 16

amorce antisens : GTG GAT CTC GGC CTC C

24. Anticorps mono ou polyclonaux ou leurs fragments, anticorps chimériques ou immunoconjugués, caractérisés en ce qu'ils sont capables de reconnaître spécifiquement un polypeptide selon l'une quelconque des revendications 1 à 4.
25. Utilisation des anticorps selon la revendication précédente, pour la purification ou la détection d'un polypeptide selon l'une quelconque des revendications 1 à 4 dans un échantillon biologique.
26. Procédé de diagnostic *in vitro* de pathologies corrélées à une expression ou une accumulation anormale de protéines SR-p70, notamment les phénomènes de cancérisation, à partir d'un prélèvement biologique, caractérisé en ce que l'on met en contact au moins un anticorps selon la revendication 24 avec ledit prélèvement biologique, dans des conditions permettant la formation éventuelle de complexes immunologiques spécifiques entre une protéine SR-p70 et le ou lesdits anticorps et en ce que l'on détecte les complexes immunologiques spécifiques éventuellement formés.
27. Kit pour le diagnostic *in vitro* d'une expression ou une accumulation anormale de protéines SR-p70 dans un prélèvement biologique et/ou pour la mesure du taux d'expression de celles-ci dans ledit prélèvement comprenant :
 - au moins un anticorps selon la revendication 24, éventuellement fixé sur un support,
 - des moyens de révélation de la formation de complexes antigènes/anticorps spécifiques entre une protéine SR-p70 et ledit anticorps et/ou des moyens de quantification de ces complexes.
28. Méthode pour le diagnostic précoce de la formation des tumeurs caractérisée en ce que l'on met en évidence dans un échantillon de sérum prélevé chez un individu des auto-anticorps dirigés contre une protéine SR-p70 selon les étapes consistant à mettre en contact un échantillon de sérum prélevé chez un individu avec un polypeptide de l'invention, éventuellement fixé sur un support, dans des conditions permettant la formation de complexes immunologiques spécifiques entre

24. Anticorps mono ou polyclonaux ou leurs fragments, anticorps chimériques ou immunoconjugués, caractérisés en ce qu'ils sont capables de reconnaître spécifiquement un polypeptide selon l'une quelconque des revendications 1 à 4.
- 5
25. Utilisation des anticorps selon la revendication précédente, pour la purification ou la détection d'un polypeptide selon l'une quelconque des revendications 1 à 4 dans un échantillon biologique.
- 10
26. Procédé de diagnostic *in vitro* de pathologies corrélées à une expression ou une accumulation anormale de protéines SR-p70, notamment les phénomènes de cancérisation, à partir d'un prélèvement biologique, caractérisé en ce que l'on met en contact au moins un anticorps selon la revendication 24 avec ledit prélèvement biologique, dans des conditions
- 15
- permettant la formation éventuelle de complexes immunologiques spécifiques entre une protéine SR-p70 et le ou lesdits anticorps et en ce que l'on détecte les complexes immunologiques spécifiques éventuellement formés.
- 20
27. Kit pour le diagnostic *in vitro* d'une expression ou une accumulation anormale de protéines SR-p70 dans un prélèvement biologique et/ou pour la mesure du taux d'expression de celles-ci dans ledit prélèvement comprenant :
- au moins un anticorps selon la revendication 24, éventuellement fixé sur un support,
- 25
- des moyens de révélation de la formation de complexes antigènes/anticorps spécifiques entre une protéine SR-p70 et ledit anticorps et/ou des moyens de quantification de ces complexes.
- 30
28. Méthode pour le diagnostic précoce de la formation des tumeurs caractérisée en ce que l'on met en évidence dans un échantillon de sérum prélevé chez un individu des auto-anticorps dirigés contre une protéine SR-p70 selon les étapes consistant à mettre en contact un échantillon de sérum prélevé chez un individu avec un polypeptide de l'invention, éventuellement fixé sur un support, dans des conditions permettant la
- 35
- formation de complexes immunologiques spécifiques entre ledit polypeptide et les auto-anticorps éventuellement présents dans l'échantillon de sérum,

et en ce que l'on détecte les complexes immunologiques spécifiques éventuellement formés.

5 29. Composition pharmaceutique comprenant, à titre de principe actif, un polypeptide selon l'une quelconque des revendications 1 à 4.

30. Composition pharmaceutique selon la revendication précédente, caractérisée en ce qu'elle comprend un polypeptide selon la revendication 2.

10

31. Composition pharmaceutique contenant un inhibiteur ou un activateur de l'activité du SR-p70.

15

20

25

30

35

ledit polypeptide et les auto-anticorps éventuellement présents dans l'échantillon de sérum, et en ce que l'on détecte les complexes immunologiques spécifiques éventuellement formés.

29. Composition pharmaceutique comprenant, à titre de principe actif, un polypeptide selon l'une quelconque des revendications 1 à 4.
30. Composition pharmaceutique selon la revendication précédente, caractérisée en ce qu'elle comprend un polypeptide selon la revendication 2.
31. Composition pharmaceutique contenant un inhibiteur ou un activateur de l'activité du SR-p70.

1 / 16

1 TGCCTCCCCGCCCGCGCACCCGCCCGAGGCCCTGTGCTCCTGCGAAGGGG 50
1GGGGCTCCGGGG 12
51 ACGCAGCGAAGCCGGGGCCCGCGCCAGGCCGGCCGGGACGGACGCCGATG 100
13 ACACTTGCGCTCCGGGCTGGAAGCGTGCTTTCCAAGACGGTGACACGCTT 62
101 CCCGGAGCTGCGACGGCTGCAGAGCGAGCTGCCCTCGGAGGCCGGTGTGA 150
63 CCCTGAGGATTGGCAGCCAGACTGCTTACGGGTAC...TGCCATGGAGG 109
151 GGAAGATGGCCAGTCCACCACCACCTCCCCGATGGGGGCACCACGTTT 200
110 AGCCGCAGTCAGATCCAGCATCGAGCCCCCTCTGAGTCAGGAAACATTT 159
201 GAGCACCTCTGGAGCTCTCTGGAACCAGACAGCACCTACTTCGACCTTCC 250
160 TCAGACCTATGGAACTACTTCCTGAAAACAAC.GTTCTGTCCCCCTTGC 208
251 CCAGTCAAGCCGGGGGAATAATGAGGTGGTGGGTGGCACGGATTCCAGCA 300
209 CGTCCCAAGCGGTGGATGATTTGATGCTCTCTCCGGATGATCTTGACAA 258
301 TGGACGTCTTCCACCTAGAGGGCATGACCACATCTGTCTATGGCCCAGTTC 350
259 TGG.....TTAACTGAAGACCCAGGTC 280
351 AATTTGCTGAGCAGCACCATGGACCAGATGAGCAGCCGCGCTGCCTCGGC 400
281 CAGATGAAGCTC.....CCAGAATGTCAGAGGCTGCTCCCCACA 319
401 CAGCCCGTACACCCCGGAGCAGCGCGCCAGCGTGCCCACCCATTACCCCT 450
320 TGGCCCCCACACCAGCAGCTCCTACACCGCGGCCCCCTGCACCAGCCCC. 368
451 ACGCACAGCCCAGCTCCACCTTCGACACCATGTGCCCCGCGCCTGTCATC 500
369CTCCTGGCCCCCTGTCATCCTCTGTC 393
501 CCCTCCAACACCGACTATCCCGGACCCACCACTTCGAGGTCACTTTCCA 550
394 CCTTCCCAGAAAACCTACCACGCGCAGCTACGGTTTCCGTCTGGGCTTCCT 443
551 GCAGTCCAGCACGGCCAAGTCAGCCACCTGGACGTACTCCCCACTCTTGA 600
444 GCATTCTGGAACAGCCAAGTCTGTGACTTGCACGTACTCCCTGACCTCA 493
601 AGAAACTCTACTGCCAGATCGCCAAGACATGCCCCATCCAGATCAAGGTG 650
494 ACAAGATGTTTGGCAGCTGGCCAAGACCTGCCCCGTGCAGCTGTGGGTT 543
651 TCCGCCCCACCGCCCCCGGGCACCGCCATCCGGGCCATGCCTGTCTACAA 700
544 GATTCCACACCCCCGCGCGGACGCGCTCCGCGCCATGGCCATCTACAA 593
701 GAAGGCGGAGCACGTGACCGACATCGTGAAGCGCTGCCCCAACCACGAGC 750
594 GCAGTCACAGCACATGACTGAGGTCTGTGAGGCGCTGCCCCCACCATGAGC 643
751 TCGGGAGGGACTTCAACGAAGGACAGTCTGCCCCAGCCAGCCACCTCATC 800
644 GCTGCTCAGACAGCGATGGA.....CTGGCCCCCTCCTCAACATCTTATC 687
801 CGTGTGGAAGGCAATAATCTCTCGCAGTATGTGGACGACCCTGTACCCGG 850
688 CGAGTGGAAGGAAATTTGCGTGTGGAGTATTCGGATGACAGAAACATTT 737
851 CAGGCAGAGCGTCGTGGTGGCCCTATGAGCCACCACAGGTGGGGACAGAAT 900
738 TCGACATAGTGTGGTGGTGGCCCTATGAGCCGCCTGAGGTGGCTCTGACT 787

FIG.1

```

901 TCACCACCATCCTGTACAAC TTCATGTGTAACAGCAGCTGTGTGGGGGGC 950
788 GTACCACCATCCACTACAAC TACATGTGTAACAGTTCCTGTCATGGGCGGC 937
951 ATGAACCGACGGCCCATCCTCATCATCACCCTGGAGACGCGGGATGG 1000
838 ATGAACCGGAGGCCATCCTCACAATTATCACACTGGAAGACTCCAGTGG 897
1001 GCAGGTGCTGGGCGCCGGTCTTCGAGGGCCGCATCTGCGCCTGTCTTG 1050
888 TAATCTACTGGGACGGAACAGCTTTGAGGTGCGAGTTTGTGCTGTCTTG 937
1051 GCCGCGACCGAAAAGCCGATGAGGACCACTACCGGGAGCAGCAGGCCTTG 1100
938 GGAGAGACCGGCGCACAGAGGAAGAGAATTTC.....G 971
1101 AATGAGAGCTCCGCCAAGAACGGGGCTGCCAGCAAGCGCCCTTCAAGCA 1150
972 CAAGAAAGGGGAGCCTTGCCACGAGCTGCCCCCTGGGAGCACTAAGCGAG 1021
1151 GAGTCCCCCTGCCGTCCCCGCCCTGGGCCC.GGGTGTGAAGAAGCGGCGG 1199
1022 CACTGCCCAACAACACCAGCTCCTCTCCCCAGCCAAAGAAGAAACCACTG 1071
1200 CACGGAGACGAGGACACGTACTACCTGCAGGTGCGAGGCCGCGAGAACTT 1249
1072 GATGGAGAATATTTCAC.....CCTTCAGATCCGCGGGCGTGAGCGCTT 1115
1250 CGAGATCCTGATGAAGCTGAAGGAGAGCCTGGAGCTGATGGAGTTGGTGC 1299
1116 CGAGATGTTCCGAGAGCTGAATGAGGCCTTGAAGTCAAGGA..... 1157
1300 CGCAGCCGCTGGTAGACTCCTATCGGCAGCAGCAGCTCCTACAGAGG 1349
1158 TGCCCAGGCTGGGAAAGAGCCAGCGG..GGAGCAGGGCTCACTCCAGCCA 1205
1350 CCGAGTCACCTACAGCCCCATCCTACGGGCCGGTCTCTCGCCCATGAA 1399
1206 CCTGAAGTCCAAGAAGGGGAATCTACCTCCCGCCATAAAAAATTCATGT 1255
1400 CAAGGTGCACGGGGGCGTGAACAAGCTGCCCTCCGTCAACCAGCTGGTGG 1449
1256 TCAAGACAGAGGGGCTGACTCAGACTGACATTC.....TCAGTTCTTG 1300
1450 GCCAGCCTCCCCGACAGCTCGGCAGCTACACCCAACCTGGGACCTGTG 1499
1301 TTCCCCACTGAGCCTCCACCCCCATCT.CTCCCTCCCTGCCATTTTG 1349
1500 GGCTCTGGGATGCTCAACAACCACGGCCACGCAGTGCCAGCCAACAGCGA 1549
1350 AGTTCTGGGTCTTTAAACCCTTGCTTGCAATAGGTGTGTGTCAGAAGCAA 1399
1550 GATGACCAGCAGCCACGGCACCCAGTCCATGGTCTCGGGTCCCACTGCA 1599
1400 A..... 1400

```

```

1 MAQSTTTSPDGGTTFEHLWSSLEPDSTYFDLPQSSRGNNNEVVGTDSSMD 50
1 .....MEEPQSDPSIEPPLS....QETFSDLWKLLPENNVLSPLPSQAVD 41
51 VFHLEGMTTSVMAQFNLLSSTMDQMSSRAASASPYTPEHAASVPTHSPYA 100
42 DLML...SPDDLAQWLTECPGPDEAPRMSEAAPHMAPTPAAPTPA..APAP 87
101 QPSSTFDTMSPAPVIPSNTDYPGPHHFEVTFQSSSTAKSATWTYSPLLKK 150
88 APSWPL.....SSSVPSQKTYHGSYGFRGLHSGTAKSVTCTYSPDLNK 132
151 LYCQIAKTCPIQIKVSAPPPPGTAIRAMPVYKKAHVTDIVKRCPNHELG 200
133 MFCQLAKTCPVQLWVDSTPPPGSRVRAMAIYKQSQHMTVEVVRRCPHHE.. 180
201 RDFNEGQSAPASHLIRVEGNNLSQYVDDPVTGRQSVVVPYEPQVGTEFT 250
181 RCSDSGLAPPQHLLIRVEGNLRVEYSDDRNTFRHSVVVPYEPPEVGS DCT 230
251 TILYNFCNSSSCVGGMNRPIILIIITLET RDGQVLGRRSFEGRICAC PGR 300
231 TIHYNMNCSSCMGGMNRPIILTIIITLEDSSGNLLGRNSFEVRVCAC PGR 280
301 DRKADEDHYREQQALNESSAKNGAASKRAFKQSPPAVPALGPGVKRRHG 350
281 DRRTEENFRKKG..EPCHELPPGSTKRALPNNTSSSPQ.....PKKKPL 323
351 DEDTYYLQVRGRENFEILMKLKESELMELVPQPLVDSYRQQQQLLQRPS 400
324 DGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEPAGSRAHSSHLKSKK 373
401 HLQPPSYGPVLSPMNKVHGQVKNKLPSVNQLVGQPPPHSSAATPNLGPVGS 450
374 GQSTSRHKKFMFKTEGPDS..... 393

```

FIG. 2


```

1 TGCCTCCCCGCCCGCGCACCCGCCCGAGGCCTGTGCTCCTGCGAAGGGG 50
  |||
1 TGCCTCCCCGCCCGCGCACCCGCCCGAGGCCTGTGCTCCTGCGAAGGGG 50
  |||
51 ACGCAGCGAAGCCGGGGCCCGCGCCAGGCCGGCCGGACGGACGCCGATG 100
  |||
51 ACGCAGCGAAGCCGGGGCCCGCGCCAGGCCGGCCGGACGGACGCCGATG 100
  |||
101 CCCGGAGCTGCGACGGCTGCAGAGCGAGCTGCCCTCGGAGGCCGGTGTGA 150
  |||
101 CCCGGAGCTGCGACGGCTGCAGAGCGAGCTGCCCTCGGAGGCCGGTGTGA 150
  |||
151 GGAAGATGGCCAGTCCACCACCACCTCCCCGATGGGGGCACCACGTTT 200
  |||
151 GGAAGATGGCCAGTCCACCACCACCTCCCCGATGGGGGCACCACGTTT 200
  |||
201 GAGCACCTCTGGAGCTCTCTGGAACCAGACAGCACCTACTTCGACCTTCC 250
  |||
201 GAGCACCTCTGGAGCTCTCTGGAACCAGACAGCACCTACTTCGACCTTCC 250
  |||
251 CCAGTCAAGCCGGGGGAATAATGAGGTGGTGGGTGGCACGGATTCCAGCA 300
  |||
251 CCAGTCAAGCCGGGGGAATAATGAGGTGGTGGGTGGCACGGATTCCAGCA 300
  |||
301 TGGACGTCTTCCACCTAGAGGGCATGACCACATCTGTCTATGGCCAGTTC 350
  |||
301 TGGACGTCTTCCACCTAGAGGGCATGACCACATCTGTCTATGGCCAGTTC 350
  |||
351 AATTTGCTGAGCAGCACCATGGACCAGATGAGCAGCCGCGCTGCCTCGGC 400
  |||
351 AATTTGCTGAGCAGCACCATGGACCAGATGAGCAGCCGCGCTGCCTCGGC 400
  |||
401 CAGCCCGTACACCCCGGAGCAGCCGCCAGCGTGCCCAACCCATTACCCCT 450
  |||
401 CAGCCCGTACACCCCGGAGCAGCCGCCAGCGTGCCCAACCCATTACCCCT 450
  |||
451 ACGCACAGCCAGCTCCACCTTCGACACCATGTGCGCCGCGCCTGTCATC 500
  |||
451 ACGCACAGCCAGCTCCACCTTCGACACCATGTGCGCCGCGCCTGTCATC 500
  |||
501 CCCTCCAACACCGACTATCCCGGACCCCAACCACTTCGAGGTCACTTTCCA 550
  |||
501 CCCTCCAACACCGACTATCCCGGACCCCAACCACTTCGAGGTCACTTTCCA 550
  |||
551 GCAGTCCAGCAGGCCAAATCAGCCACCTGGACGTACTCCCCACTCTTGA 600
  |||
551 GCAGTCCAGCAGGCCAAATCAGCCACCTGGACGTACTCCCCACTCTTGA 600
  |||
601 AGAAACTCTACTGCCAGATCGCCAAGACATGCCCCATCCAGATCAAGGTG 650
  |||
601 AGAAACTCTACTGCCAGATCGCCAAGACATGCCCCATCCAGATCAAGGTG 650
  |||
651 TCCGCCCCACCGCCCCCGGCCACCGCCATCCGGGCCATGCCTGTCTACAA 700
  |||
651 TCCGCCCCACCGCCCCCGGCCACCGCCATCCGGGCCATGCCTGTCTACAA 700
  |||
701 GAAGGCGGAGCAGCTGACTGACATCGTGAAGCGCTGCCCCAACCACGAGC 750
  |||
701 GAAGGCGGAGCAGCTGACTGACATCGTGAAGCGCTGCCCCAACCACGAGC 750
  |||
751 TCGGGAGGGACTTCAACGAAGGACAGTCTGCCCCAGCCAGCCACCTCATC 800
  |||
751 TCGGGAGGGACTTCAACGAAGGACAGTCTGCCCCAGCCAGCCACCTCATC 800
  |||
801 CGTGTGGAAGGCAATAATCTCTCGCAGTATGTGGACGACCCTGTCAACCG 850
  |||
801 CGTGTGGAAGGCAATAATCTCTCGCAGTATGTGGACGACCCTGTCAACCG 850
  |||
851 CAGGCAGAGCGTCGTGGTGGCCCTATGAGCCACCACAGGTGGGGACAGAAT 900
  |||
851 CAGGCAGAGCGTCGTGGTGGCCCTATGAGCCACCACAGGTGGGGACAGAAT 900
  |||

```

FIG. 3

```

901 TCACCACCATCCTGTACAACCTTCATGTGTAACAGCAGCTGTGTGGGGGGC 950
|||||
901 TCACCACCATCCTGTACAACCTTCATGTGTAACAGCAGCTGTGTGGGGGGC 950
|||||
951 ATGAACCGACGSCCCATCCTCATCATCATCACCCCTGGAGACGCGGGATGG 1000
|||||
951 ATGAACCGACGSCCCATCCTCATCATCATCACCCCTGGAGACGCGGGATGG 1000
|||||
1001 GCAGGTGCTGGGCCGCGGTCCTTCGAGGGCCGCATCTGCGCCTGTCTCTG 1050
|||||
1001 GCAGGTGCTGGGCCGCGGTCCTTCGAGGGCCGCATCTGCGCCTGTCTCTG 1050
|||||
1051 GCCGCGACCGAAAAGCCGATGAGGACCACTACCGGGAGCAGCAGGCCTTG 1100
|||||
1051 GCCGCGACCGAAAAGCCGATGAGGACCACTACCGGGAGCAGCAGGCCTTG 1100
|||||
1101 AATGAGAGCTCCGCCAAGAACGGGGCTGCCAGCAAGCGCGCCTTCAAGCA 1150
|||||
1101 AATGAGAGCTCCGCCAAGAACGGGGCTGCCAGCAAGCGCGCCTTCAAGCA 1150
|||||
1151 GAGTCCCCCTGCCGTCCCCGCCCTGGGCCCGGGTGTGAAGAAGCGCGGC 1200
|||||
1151 GAGTCCCCCTGCCGTCCCCGCCCTGGGCCCGGGTGTGAAGAAGCGCGGC 1200
|||||
1201 ACGGAGACGAGGACACGTACTACCTGCAGGTGCGAGGCCGCGAGAACTTC 1250
|||||
1201 ACGGAGACGAGGACACGTACTACCTGCAGGTGCGAGGCCGCGAGAACTTC 1250
|||||
1251 GAGATCCTGATGAAGCTGAAGGAGAGCCTGGAGCTGATGGAGTTGGTGCC 1300
|||||
1251 GAGATCCTGATGAAGCTGAAGGAGAGCCTGGAGCTGATGGAGTTGGTGCC 1300
|||||
1301 GCAGCCGCTGGTAGACTCCTATCGGCAGCAGCAGCAGCTCCTACAGAGGC 1350
|||||
1301 GCAGCCGCTGGTAGACTCCTATCGGCAGCAGCAGCAGCTCCTACAGAGGC 1350
|||||
1351 CGAGTCACCTACAGCCCCCATCTACGGGCCGGTCCTCTCGCCCATGAAC 1400
|||||
1351 CGAGTCACCTACAGCCCCCATCTACGGGCCGGTCCTCTCGCCCATGAAC 1400
|||||
1401 AAGGTGCACGGGGGCGTGAACAAGCTGCCCTCCGTCAACCAGCTGGTGGG 1450
|||||
1401 AAGGTGCACGGGGGCGTGAACAAGCTGCCCTCCGTCAACCAGCTGGTGGG 1450
|||||
1451 CCAGCCTCCCCCGCACAGCTCGGCAGCTACACCCAACCTGGGACCTGTGG 1500
|||||
1451 CCAGCCTCCCCCGCACAGCTCGGCAGCTACACCCAACCTGGGACCTGTGG 1500
|||||
1501 GCTCTGGGATGCTCAACAACCACGGCCACGCAGTGCCAGCCAACAGCGAG 1550
|||||
1501 GCTCTGGGATGCTCAACAACCACGGCCACGCAGTGCCAGCCAACAGCGAG 1550
|||||
1551 ATGACCAGCAGCCACGGCACCCAGTCCATGGTCTCGGGGTCCCAGTGCAC 1600
|||||
1551 ATGACCAGCAGCCACGGCACCCAGTCCATGGTCTCGGGGTCCCAGTGCAC 1600
|||||
1601 TCCGCCACCCCTTACCACGCCGACCCAGCCTCGTCAGTTTTTTAACAG 1650
|||||
1601 TCCGCCACCCCTTACCACGCCGACCCAGCCTCGTC..... 1637
|||||

1701 AGCATTTACCACCTGCAGAACCTGACCATCGAGGACCTGGGGGCCCTGAA 1750
|||||
1638 .....AGGACCTGGGGGCCCTGAA 1656
|||||
1751 GATCCCCGAGCAGTATCGCATGACCATCTGGCGGGGCTGCAGGACCTGA 1800
|||||

```

FIG. 3
suite

```

1657 GATCCCCGAGCAGTATCGCATGACCATCTGGCGGGGCCTGCAGGACCTGA 1706
1801 AGCAGGGCCACGACTACGGCGCCGCCGCGCAGCAGCTGCTCCGCTCCAGC 1850
      |||
1707 AGCAGGGCCACGACTACGGCGCCGCCGCGCAGCAGCTGCTCCGCTCCAGC 1756
      |||
1851 AACGCGGGCCGCCATTTCCATCGGGGGCTCCGGGGAGCTGCAGCGCCAGCG 1900
      |||
1757 AACGCGGGCCGCCATTTCCATCGGGGGCTCCGGGGAGCTGCAGCGCCAGCG 1806
      |||
1901 GGTCATGGAGGCCGTGCACTTCCGCGTGCGCCACACCATCACCATCCCCA 1950
      |||
1807 GGTCATGGAGGCCGTGCACTTCCGCGTGCGCCACACCATCACCATCCCCA 1856
      |||
1951 ACCGCGGGCGGGCCCGGGCGCCGGCCCCGACGAGTGGGCGGACTTCGGCTTC 2000
      |||
1857 ACCGCGGGCGGGCCCGGGCGCCGGCCCCGACGAGTGGGCGGACTTCGGCTTC 1906
      |||
2001 GACCTGCCCCGACTGCAAGGCCCGCAAGCAGCCCATCAAGGAGGAGTTCAC 2050
      |||
1907 GACCTGCCCCGACTGCAAGGCCCGCAAGCAGCCCATCAAGGAGGAGTTCAC 1956
      |||
2051 GGAGGCCGAGATCCACTGAGGGGCGGGGCCAGCCAGAGCCTGTGCCACC 2100
      |||
1957 GGAGGCCGAGATCCACTGAGGGGCGGGGCCAGCCAGAGCCTGTGCCACC 2006
      |||
2101 GCCCAGAGACCCAGGCCGCCTCGCTCTC 2128
      |||
2007 GCCCAGAGACCCAGGCCGCCTCGCTCTC 2034

```

FIG.3
suite

1	TGCCTCCCCGCGCGCACCCGCCCCGAGGCCTGTGCTCCTGCGAAGGGGACGCAGCGAA	60
61	GCCGGGGCCCGGCCAGGCCGCGCGGGACGGACGCCGATGCCCGGAGCTGCGACGGCTGC	120
121	AGAGCGAGCTGCCCTCGGAGGCCGGTGTGAGGAAGATGGCCAGTCCACCACCACCTCCC	180
-10	M A Q S T T T S P	9
181	CCGATGGGGGACACACGTTTGTAGCACCTCTGGAGCTCTCTGGAACCAGACAGCACCTACT	240
10	D G G T T F E H L W S S L E P D S T Y F	29
241	TCGACCTTCCCCAGTCAAGCCGGGGGAATAATGAGGTGGTGGGTGGCAGGATTCCAGCA	300
30	D L P Q S S R G N N E V V G G T D S S M	49
301	TGGACGTCTTCCACCTAGAGGGCATGACCACATCTGTCATGGCCAGTTCAATTGCTGA	360
50	D V F H L E G M T T S V M A Q F N L L S	69
361	GCAGACCATGGACCAGATGAGCAGCCGCGCTGCCTCGGCCAGCCCGTACACCCCGGAGC	420
70	S T M D Q M S S R A A S A S P Y T P E H	89
421	ACGCGCCAGCGTGCCCAACCCATTACCCCTACGCACAGCCAGCTCCACCTTCGACACCA	480
90	A A S V P T H S P Y A Q P S S T F D T M	109
481	TGTCGCGCGCGCTGTCTATCCCTCCAACACCGACTATCCCGGACCCACCACCTTCGAGG	540
110	S P A P V I P S N T D Y P G P H H F E V	129
541	TCACTTTCCAGCAGTCCAGCAGCGCCAAGTCAGCCACCTGGACGTACTCCCACTCTGA	600
130	T F Q Q S S T A K S A T W T Y S P L L K	149
601	AGAACTCTACTGCCAGATCGCCAAGACATGCCCCATCCAGATCAAGGTGTCCGCCCCAC	660
150	K L Y C Q I A K T C P I Q I K V S A P P	169
661	CGCCCCCGGGCACC GCCATCCGGCCATGCCTGTCTACAAGAAGGCGGAGCAGGTGACCG	720
170	P P G T A I R A M P V Y K K A E H V T D	189
721	ACATCGTGAAGCGCTGCCCAACCCAGAGCTCGGGAGGGACTTCAACGAAGGACAGTCTG	780
190	I V K R C P N H E L G R D F N E G Q S A	209
781	CCCCAGCCAGCCACCTCATCCGTGTGGAAGGCAATAATCTCTCGAGTATGTGGACGACC	840
210	P A S H L I R V E G N N L S Q Y V D D P	229
841	CTGTCAACCGGCAGGCAGAGCGTCTGGTGCCTATGAGCCACCACAGGTGGGGACAGAAT	900
230	V T G R Q S V V V P Y E P P Q V G T E F	249
901	TCACCACCATCCTGTACAACCTCATGTGTGTAACAGCAGCTGTGTGGGGGCGTGAACCGAC	960
250	T T I L Y N F M C N S S C V G G M N R R	269
961	GGCCCCCTCATCATCATCACCTGGAGACGCGGGATGGGCAGGTGCTGGGCGCGCGGT	1020
270	P I L I I I T L E T R D G Q V L G R R S	289
1021	CCTTCGAGGGCGCATCTGCGCTGTCTGCTGGCCGCGACCGAAAAGCCGATGAGGACCACT	1080
290	F E G R I C A C P G R D R K A D E D H Y	309
1081	ACCGGGAGCAGCAGGCCTTGAATGAGAGCTCCGCCAAGAACGGGGCTGCCAGCAAGCCGC	1140
310	R E Q Q A L N E S S A K N G A A S K R A	329
1141	CCTTGAAGCAGAGTCCCCCTGCCGTCCCGCCCTGGGCCCCGGGTGTGAAGAAGCGCGGC	1200
330	F K Q S P P A V P A L G P G V K K R R H	349
1201	ACGGAGACGAGGACACGTACTACCTGCAAGGTGCGAGGCCGCGAGAACTTCGAGATCCTGA	1260
350	G D E D T Y Y L Q V R G R E N F E I L M	369
1261	TGAAGCTGAAGGAGAGCCTGGAGCTGATGGAGTTGGTGCCGACGCCGCTGGTAGACTCCT	1320
370	K L K E S L E L M E L V P Q P L V D S Y	389
1321	ATCGGCAGCAGCAGCAGCTCTACAGAGGCCGAGTCACCTACAGCCCCCATCCTACGGGC	1380
390	R Q Q Q Q L L Q R P S H L Q P P S Y G P	409
1381	CGGTCTCTCGCCCATGAACAAGGTGCACGGGGCGTGAACAAGCTGCCCTCCGTCAACC	1440
410	V L S P M N K V H G G V N K L P S V N Q	429
1441	AGCTGGTGGGCCAGCCTCCCCCGCACAGCTCGGCAGCTACACCCAACCTGGGACCTGTGG	1500
430	L V G Q P P P H S S A A T P N L G P V G	449
1501	GCTCTGGGATGCTCAACAACACGGCCACGCAGTGCCAGCCAACAGCGAGATGACCAGCA	1560
450	S G M L N N H G H A V P A N S E M T S S	469
1561	GCCACGGCACCCAGTCCATGCTCGGGGTCCCACTGCCTCCGCCACCCCCCTACCAAG	1620
470	H G T Q S M V S G S H C T P P P P Y H A	489
1621	CCGACCCAGCCTCGTCACTTTTAAACAGGATGGGGTGTCCAAACTGCATCGAGTATT	1680
490	D P S L V S F L T G L G C P N C I E Y F	509

FIG.4

1681	TCACGTCCCAGGGTTACAGAGCATT TACCACT GCAGAACCTGACCATCGAGGACCTGG	1740
510	T S Q G L Q S I Y H L Q N L T I E D L G	529
1741	GGGCCCTGAAGATCCCCGAGCAGT ATCCAT GACCATCTGGCGGGGCTGCAGGACCTGA	1800
530	A L K I P E Q Y R N T I W R G L Q D L K	549
1801	AGCAGGGCCACGACTACGGCGCCG CCGCGCAGCAG CTGCTCCGCTCCAGCAACGCGGCCG	1860
550	Q G H D Y G A A A Q Q L L R S S N A A A	569
1861	CCATTTCCATCGGCGGCTCCGGGG AGCTGCAGCGCC CAGCGGGTCATGGAGGCCGTGCACT	1920
570	I S I G G S G E L Q R Q R V M E A V H F	589
1921	TCCGCGTGCGCCACACCATCACC ATCCCCAAC CGCGCGCGCCCGCGCGCCCGCCGACG	1980
590	R V R H T I T I P R G G P G A G P D E	609
1981	AGTGGCGGACTTCGGCTTCGAC CTGCCCCG ACTGCAAGGCCCGCAAGCAGCCCATCAAGG	2040
610	W A D F G F D L P D C K A R K Q P I K E	629
2041	AGGAGTTCACGGAGGCCGAGATCC ACTGAGGG CCCGGGCCAGCCAGAGCCTGTGCCACC	2100
630	E F T E A E I H *	649
2101	GCCCAGAGACCCAGGCCGCTCGCT CTCTCT CTCTGTGTCCAAACTGCCTCCGGAGGCAG	2160
2161	GGCCTCCAGGCTGTGCCCGGGG AAAGCAAG CTCCGGCCCATGCCCCGGCACCTCACC	2220
2221	CCCCAGGAGAGGCCAGCCACCA AGCCGCTG CGGACAGCCTGAGTCACCTGCAGAACC	2280
2281	TTCTGGAGCTGCCCTAATGCTGGG CTTGGCGG CAGGGGCCCGGCCACTCTCAGCCCTGC	2340
2341	CACTGCCCGGCGTGCTCCATGGCAGG CGTGGGT GGGGACCGCAGTGTCACTCCGACCTC	2400
2401	CAGGCCCTCATCCTAGAGACTCTGT CATCTG CCGATCAAGCAAGGTCCTTCCAGAGGAAAG	2460
2461	AATCCTCTTCGCTGGTGGACTGCC AAAAAG TATTTTGCACATCTTTTGGTTCTGGAGAG	2520
2521	TGGTGAGCAGCCAAGCGACTGTGT CTGAAAC ACCGTGCATTTTCAGGGAATGTCCCTAAC	2580
2581	GGGCTGGGACTCTCTCTGCTGG ACTTGGG AGTGGCCTTTGCCCCAGCACACTGTATT	2640
2641	TGCGGGACCGCCTCCTTCTGCCC CTAAC ACCAACCAAGTGTGCTGAAATTGGAGAAA	2700
2701	ACTGGGGAAGGCGCAACCCCTCC AGGTG CGGGAAGCATCTGGTACCGCCTCGGCCAGTG	2760
2761	CCCCTCAGCCTGGCCACAGTCAC CTCTCT TGGGGAACCCCTGGGCAGAAAGGGACAGCCT	2820
2821	GTCCTTAGAGGACCGGAAATTGT CAATATT TGATAAAATGATACCCTTTTCTAC	2874

FIG.4

suite

1 TGCCTCCCCGCCCGCGCACCCGCCCGAGGCCTGTGCTCCTGCGAAGGGGACGCAGCGAA 60
 61 GCCGGGGCCCGCGCCAGGCCGGCGGGACGGACGCCGATGCCCGGAGCTGCGACGGCTGC 120
 121 AGAGCGAGCTGCCCTCGGAGGCCGCTGTGAGGAAGATGGCCCCAGTCCACCACCACCTCCC 180
 -10 M A Q S T T T S P 9
 181 CCGATGGGGGCACCACGTTTGTAGACCTCTGGAGCTCTCTGGAACCAGACAGCACCTACT 240
 10 D G G T T F E H L W S S L E P D S T Y F 29
 241 TCGACCTTCCCCAGTCAAGCCGGGGGAATAATGAGGTGGTGGGTGGCAGGATTCCAGCA 300
 30 D L P Q S S R G N N E V V G G T D S S M 49
 301 TGGACGTCTTCCACCTAGAGGGCATGACCACATCTGTCATGGCCCAGTTCAATTTGTGA 360
 50 D V F H L E G M T T S V M A Q F N L L S 69
 361 GCAGCACCATGGACCAGATGAGCAGCCGCGCTGCCTCGGCCAGCCGTACACCCCGGAGC 420
 70 S T M D Q M S S R A A S A S P Y T P E H 89
 421 ACGCCGCCAGCGTGCCACCCATTACCCCTACGCACAGCCAGCTCCACCTTCGACACCA 480
 90 A A S V P T H S P Y A Q P S S T F D T M 109
 481 TGTGCGCCCGCGCTGTCTATCCCTTCAACACCGACTATCCCGGACCCACCACTTCGAG 540
 110 S P A P V I P S N T D Y P G P H H F E V 129
 541 TCACTTTCCAGCAGTCCAGCAGGCCAAGTCAGCCACCTGGACGTACTCCCCACTCTTGA 600
 130 T F Q Q S S T A K S A T W T Y S P L L K 149
 601 AGAAACTCTACTGCCAGATCGCCAAGACATGCCCATCCAGATCAAGGTGTCCGCCCCAC 660
 150 K L Y C Q I A K T C P I Q I K V S A P P 169
 661 CGCCCCCGGGCACCGCCATCCGGGCCATGCTCTCTACAAGAAGCGGAGCACGTGACCG 720
 170 P P G T A I R A M P V Y K K A E H V T D 189
 721 ACATCGTGAAGCGTGCCCAACCAGAGCTCGGGAGGGACTTCAACGAAGGACAGTCTG 780
 190 I V K R C P N H E L G R D F N E G Q S A 209
 781 CCCCAGCCAGCCACCTCATCCGTGTGAAGGCAATCTCTCGCAGTATGTGGACGACC 840
 210 P A S H L I R V E G N N L S Q Y V D D P 229
 841 CTGTACCCGGCAGGCAGAGCGTCGTGGTGCCTATGAGCCACCACAGGTGGGGACAGAAT 900
 230 V T G R Q S V V V P Y E P P Q V G T E F 249
 901 TCACCACCATCCTGTACAACCTTCATGTGTAAACAGCAGCTGTGTGGGGGGCATGAACCGAC 960
 250 T T I L Y N F M C N S S C V G G M N R R 269
 961 GGCCCATCCTCATCATCATCACCTGGAGACCGGGATGGGCAGGTGTGTGGCCGCCCGGT 1020
 270 P I L I I I T L E T R D G Q V L G R R S 289
 1021 CCTTCGAGGGGCCGATCTGCGCCTGTCTGCGCCGCGACCGAAAAGCCGATGAGGACCACT 1080
 290 F E G R I C A C P G R D R K A D E D H Y 309
 1081 ACCGGGAGCAGCAGGCCTTGAATGAGAGCTCCGCCAAGAAGCGGGCTGCCAGCAAGCGCG 1140
 310 R E Q Q A L N E S S A K N G A A S K R A 329
 1141 CCTTCAAGCAGAGTCCCCCTGCCGTCCCCCGCTGGGCCCCGGGTGTGAAGAAGCGGCGGC 1200
 330 F K Q S P P A V P A L G P G V K K R R H 349
 1201 ACGGAGACGAGGACACGTACTACCTGCAGGTGCGAGGCCGCGAGAACTTCGAGATCCTGA 1260
 350 G D E D T Y Y L Q V R G R E N F E I L M 369
 1261 TGAAGCTGAAGGAGAGCCTGGAGTGTGAGTTGGTGCCGACGCGCTGGTAGACTCCT 1320
 370 K L K E S L E L M E L V P Q P L V D S Y 389
 1321 ATCGGCAGCAGCAGCAGCTCCTACAGAGGCCGAGTACCTACAGCCCCATCCTACGGGC 1380
 390 -R Q Q Q Q L L Q R P S H L Q P P S Y G P 409
 1381 CGGTCTCTCTCGCCCATGAACAAGGTGCACGGGGCGTGAACAAGCTGCGCTCCGTCAACC 1440
 410 V L S P M N K V H G G V N K L P S V N Q 429
 1441 ACTGGTGGGCCAGCCTCCCCGSCACAGCTCGGCAGCTACACCCAACCTGGGACCTGTGG 1500
 430 L V G Q P P P H S S A A T P N L G P V G 449
 1501 GCTCTGGGATGCTCAACAACCACGCCACGCAGTGCCAGCCAACAGCGAGATGACCAGCA 1560
 450 S G M L N N H G H A V P A N S E M T S S 469
 1561 GCCACGGCACCCAGTCCATGGTCTCGGGGTCCCACTGCACCTCCGCCACCCCTACCAG 1620
 470 H G T Q S M V S G S H C T P P P P Y H A 489
 1621 CCGACCCAGCCTCGTCAGGACCTGGGGGCCCTGAAGATCCCCGAGCAGTATCGCATGAC 1680
 490 D P S L V R T W G P 509
 1681 CATCTGGCGGGGCTGCAGGACCTGAAGCAGGCGCCACGACTACGGCGCCCGCGCGAGCA 1740
 1741 GCTGCTCCGCTCCAGCAACGCGGCGCCATTTCCATCGGCGGCTCCGGGGAGCTGCAGCG 1800
 1801 CCAGCGGGTTCATGGAGGCGGTGCACTTCCGCGTGCGCCACACCATCACCATCCCCAACCG 1860
 1861 CGCGGCCCCCGCGCGGCCCCGACGAGTGGGCGGACTTCGGCTTCGACCTGCCCGACTG 1920
 1921 CAAGGCCCCGCAAGCAGCCCATCAAGGAAGAGTTACGGAGGCCGAGATCCACTGAGGGGC 1980
 1981 CGGGCCAGCCAGAGCCTGTGCCACCGCCAGAGACCCAGGCCGCGCTCGCTCTC 2034

FIG. 5

1 GCGAGCTGCCCTCGGAGGCCGGCGTGGGGAAGATGGCCCCAGTCCACCGCCACCTCCCCTG 60
 -9 M A Q S T A T S P D 10
 61 ATGGGGGACCACGTTTGTAGCACCTCTGGAGCTCTCTGGAACCAGACAGCACCTACTTCG 120
 11 G G T T F E H L W S S L E P D S T Y F D 30
 121 ACCTTCCCCAGTCAAGCCGGGGAATAATGAGGTGGTGGGcGGAACGGATTCCAGCATGG 180
 31 L P Q S S R G N N E V V G G T D S S M D 50
 181 ACGTCTTCCACCTGGAGGGCATGACTACATCTGTTCATGGCCCCAGTTCAATCTGCTGAGCA 240
 51 V F H L E G M T T S V M A Q F N L L S S 70
 241 GCACCATGGACCAGATGAGCAGCCGCGCGCTCGGCCAGCCCCACCCAGAGCAGC 300
 71 T M D Q M S S R A A S A S P Y T P E H A 90
 301 CCGCCAGCGTGcCCACCCAcTCGCCCTACGCACAACCCAGCTCCACCTTCGACACCATGT 360
 91 A S V P T H S P Y A Q P S S T F D T M S 110
 361 CGCCGGCGCCTGTTCATCCCTCCAACACCGACTACCCCGACCCCACTTTGAGGTCA 420
 111 P A P V I P S N T D Y P G P H H F E V T 130
 421 CTTTCCAGCAGTCCAGCAGCGGCAAGTCAAGCACTGGACGTACTCCCGCTCTTGAAGA 480
 131 F Q Q S S T A K S A T W T Y S P L L K K 150
 481 AACTCTACTGCCAGATCGCCAAGACATGCCCATCCAGATCAAGGTGTCCACCCCGCCAC 540
 151 L Y C Q I A K T C P I Q I K V S T P P P 170
 541 CCCCAGGCATGCCATCGGGCCATGCCTGTTTACAAGAAAGCGGagcACGTGACCGGACG 600
 171 P G T A I R A M P V Y K K A E H V T D V 190
 601 TCGTGAACCGCTGCCCAACACGAGCTCGGGAGGGACTTCAACGAAGGACAGTCTGCTC 660
 191 V K R C P N H E L G R D F N E G Q S A P 210
 661 CAGcCAGCCACCTCATCCGCGTGAAGGCAATAATCTCTCGCAGTATGTGGATGACCCTG 720
 211 A S H L I R V E G N N L S Q Y V D D P V 230
 721 TCACCGGCAGGCAGAGCTCGTGGTGCCTATGAGCCACCACAGGTGGGGACGGAATTCA 780
 231 T G R Q S V V V P Y E P P Q V G T E F T 250
 781 CCACCATCTGTACAACCTTCATGTGTAAACAGCAGCTGTGTAGGGGGCATGAACCGGGC 840
 251 T I L Y N F M C N S C V G G M N R R P 270
 841 CCATCTCATCATCACCTGGAGATGCGGGATGGGCAGGTGCTGGGCGCGCGGTCTCT 900
 271 I L I I I T L E M R D G Q V L G R R S F 290
 901 TTGAGGGCCGCATCTGCGCTGTCTGGCCGCGACCGAAAAGCTGATGAGGACCACTACC 960
 291 E G R I C A C P G R D R K A D E D H Y R 310
 961 GGGAGCAGCAGGCCCTGAACGAGAGCTCCGCCAAGAACGGGGCCGCCAGCAAGCGTGCCT 1020
 311 E Q Q A L N E S S A K N G A A S K R A F 330
 1021 TCAAGCAGAGCCCCCTGCCGTcCCGCCCTTGGTGCCGGTGTGAAGAAGCGGCGCATG 1080
 331 K Q S P P A V P A L G A G V K K R R H G 350
 1081 GAGACGAGGACACGTACTACCTTCAGGTGCGAGGCCGGGAGAACTTTGAGATCTCTGATGA 1140
 351 D E D T Y Y L Q V R G R E N F E I L M K 370
 1141 AGTGAAGAGAGCCTGGAGCTGAGTGTGGTGGCGCAGCCACTGGTGGACTCCTATC 1200
 371 L K E S L E L M E L V P Q P L V D S Y R 390
 1201 GGCAGCAGCAGCAGCTCCTACAGAGGCCGAGTCACTACAGCCCCCGTCTACGGGCCGG 1260
 391 Q Q Q Q L L Q R P S H L Q P P S V 410
 1261 TCCTCTCGCCCATGAACAAGGTGCACGGGGCATGAACAAGCTGCCCTCCGTCAACCAGC 1320
 411 L S P M N K V H G G M N K L P S V N Q L 430
 1321 TGGTGGGCCAGCTCCCCCGCACAGTTCGGCAGCTACACCCAACCTGGGGCCCGGGCC 1380
 431 V G Q P P P H S S A A T P N L G P V G P 450
 1381 CCGGGATGcTCAACAACCATGGcCACGCAGTGCAGCCAacGGCGAGATGAGCAGAGCC 1440
 451 G M L N N H G H A V P A N G E M S S H 470
 1441 ACAGCGCCAGTCCATGGTCTCGGGGTCCCACTGcACTCCGCCACCCCCCTACCACGCCG 1500
 471 S A Q S M V S G S H C T P P P P Y H A D 490
 1501 ACCCCAGCCTCGTCAGTTTTTTAACTAGGATTGGGGTGTCCAACTGCATCGAGTATTCA 1560
 491 P S L V S F L T G L G C P N C I E Y F T 510
 1561 CCTCCCAAGGGTTACAGAGCATTTACCACCTgcAGAACcTGACcATTGAGGACCTGGGGG 1620
 511 S Q G L Q S I Y H L Q N L T I E D L G A 530
 1621 CCCTGAAGATCCCCGAGCAGTACCGCATGACCATCTGGCGGGGCGCTGCAGGACCTGAAGC 1680
 531 L K I P E Q Y R M T I W R G L Q D L K Q 550
 1681 AGGGCCACGACTACAGCACCGCGCAGCAGCTGCTCCGCTCTAGCAACCGGGCCACCATCT 1740
 551 G H D Y S T A Q Q L L R S S N A A T I S 570
 1741 CCATCGGGCGGCTCAGGGGAACCTGCAGCGCCAGCGGTCATGGAGGCGCTGCACCTTCCGCG 1800
 571 I G G S G E L Q R Q R V M E A V H F R V 590
 1801 TCGCCACACCATCACCATCCCCAACCGCGCGCGCCAGGCGGGCGGCTGACGAGTGGV 1860
 591 R H T I T I P N R G G P G G G P D E W A 610
 1861 CGGACTTCGGCTTCGACCTGCCCGACTGCAAGGCCCGCAAGCAGCCCATCAAGGAGGAGT 1920
 611 D F G F D I P D C K A R K Q P I K E E F 630
 1921 TCACGGAGGCGAGATCCACTGAGGGCTCGCCTGGCTGCAGCCTGGGcACCGCCCA 1980
 631 T E A E I H 650
 1981 GACCCAAGCTGCCTCCCTCTCTCTGTGTSTCCAAAACCTGCCTCAGGAGGCAGGACC 2040
 2041 TTCGGGCTGTGCCGGGGGAAAGGCAAGTCCGGCCATCCCCAGGACCTCACAGGCCCC 2100
 2101 AGGAAAGGCCAGCCACCGAAGCGCGCTGTGGACAGCCTGAGTCACCTGCAGAACC 2155

FIG. 6

1	TGATCTCCCTGTGGCCTGCAGGGGACTGAGCCAGGGAGTAGATGCCCTGAGACCCCAAGG	60
61	GACACCCAAGGAAACCTTGCTGGCTTTGAGAAAGGGATCGTCTCTCTCCTGCCCAAGAGA	120
121	AGCATGTGTATGGGCCCCTGTGTATGAATCCTTGGGGCAGGCCAGTTCAATTTGCTCAGC	180
0	M C M G P V Y E S L G Q A Q F N L L S	19
181	AGTGCCATGGACCAGATGGGCAGCCGTGCGGCCCCGGCGAGCCCTACACCCCGGAGCAC	240
20	S A M D Q M G S R A A P A S P Y T P E H	39
241	GCCGCCAGCGCGCCCACTCGCCCTACGCGCAGCCCAGCTCCACCTTCGACACCATG	300
40	A A S A P T H S P Y A Q P S S T F D T M	59
301	TCTCCGGCGCTGTATCCCTTCCAATACCGACTACCCCGGCCCCCACCCTCGAGGTC	360
60	S P A P V I P S N T D Y P G P H H F E V	79
361	ACCTTCCAGCAGTCGAGCACTGCCAAGTCGGCCACCTGGACATACTCCCCACTCTGAAG	420
80	T F Q Q S S T A K S A T W T Y S P L L K	99
421	AAGTTGTACTGTCTAGATTGCTAAGACATGCCCCATCCAGATCAAAGTGTCACACCCACA	480
100	K L Y C Q I A K T C P I Q I K V S T P P	119
481	CCCCGGGCACGGCCATCCGGGCCATGCTGTCTACAAGAAGGCAGAGCATGTGACCGAC	540
120	P P G T A I R A M P V Y K K A E H V T D	139
541	ATTGTTAAGCGCTGCCCAACCACGAGCTTGAAGGGGACTTCAATGAAGGACAGTCTGCC	600
140	I V K R C P N H E L G R D F N E G Q S A	159
601	CCGGCTAGCCACCTCATCCGTGTAGAAGGCAACAACCTCGCCAGTACGTGAGACCCCT	660
160	P A S H L I R V E G N N L A Q Y V D D P	179
661	GTCACCGGAAGGCAGAGTGTGGTTGTGCCGTATGAACCCCCACAGGTGGGAACAGAAATT	720
180	V T G R Q S V V V Y E P P Q V G T E F	199
721	ACCACCATCCTGTACAATTCATGTGTAAACAGCAGCTGTGTGGGGGCATGAATCGGAGG	780
200	T T I L Y N F M C N S S C V G G M N R R	219
781	CCCATCCTTGTCATCATCACCTGGAGACCCGGGATGGACAGGTCTGGGCCCGCCGCT	840
220	P I L V I I T L E T R D G Q V L G R R S	239
841	TTGAGGGTCGCATCTGTGCCTGTCTGGCCGTGACCGCAAAGCTGATGAAGACCATTAC	900
240	F E G R I C A C P G R D R K A D E D H Y	259
901	CGGGAGCAACAGGCTCTGAATGAAAGTACCACCAAAATGGAGCTGCCAGCAAACTGCA	960
260	R E Q Q A L N E S T T K N G A A S K R A	279
961	TTCAAGCAGAGCCCCCTGCCATCCCTGCCCTGGGTACCAACGTGAAGAAGAGAGCCAC	1020
280	F K Q S P P A I P A L G T N V K K R H	299
1021	GGGGACGAGGACATGTTCTACATGCACGTGCGAGGCCGGGAGAACTTTGAGATCTTGATG	1080
300	G D E D M F Y M H V R G R E N F E I L M	319
1081	AAAGTCAAGGAGAGCCTAGAACTGTAGGAGCTTGTGCCCCAGCCTTTGGTTGACTCCTAT	1140
320	K V K E S L E L M E L V P Q P L V D S Y	339
1141	CGACAGCAGCAGCAGCAGCAGCTCCTACAGAGGCCGAGTCACCTGCAGCCTCCATCCTAT	1200
340	R Q Q Q Q L L Q R P S H L Q P P S Y	359
1201	GGGCCCCGTGCTCTCCCCAATGAACAAGGTACACGGTGCTGTCACAACTGCCCTCCGTC	1260
360	G P V L S P M N K V H G G V N K L P S V	379
1261	AACCAGCTGGTGGGCCAGCCTCCCCCGCAGCAGCTCAGCAGCTGGGCCCAACCTGGGGCC	1320
380	N Q L V G Q P P P H S S A A G P N L Q D	399
1321	ATGGGCTCCGGGATGCTCAACAGCCACGGCCACAGCATGCCGGCCAATGGTGAGATGAAT	1380
400	M G S G M L N S H G H S M P A N G E M N	419
1381	GGAGGCCACAGCTCCCAAGCCTGCTTGGGAGTCCCACTGcACCCCGCCACCCCTAT	1440
420	G G H S S Q T M V S G S H C T P P P P Y	439
1441	CATGCAGACCCAGCCTCCTCAGTTTTTTGACAGGGTTGGGGTGTCAAACTGCATCGAG	1500
440	H A D P S L V S F L T G L G C P N C I E	459
1501	TGCTTCACTTCCCAAGGGTTGCAGAGCATCTACCACCTGCAGAACCTTACCATCGAGGAC	1560
460	C F T S Q G L Q S I Y H L Q N L T I E D	479
1561	CTTGGGGCTCTGAAGGTCCCTGACCAGTACCGTATGACCATCTGGAGGGGCCCTACAGGAC	1620
480	L G A L K V P D Q Y R M T I W R G L Q D	499
1621	CTGAAGCAGAGCCATGACTGCGGCCAGCAACTGCTACGCTCCAGCAGCAACGCCGCCACC	1680
500	L K Q S H D C G Q Q L L R S S S N A A T	519
1681	ATCTCCATCGGCGGCTCTGGCGAGCTGCAGCGGCGGCTCATGGAAGCCGTGCATTTC	1740
520	I S I G G S G E L Q R Q R V M E A V H F	539
1741	CGTGTGGGCCACACCATCAAAATCCCCAACCGTGGAGGGCGCAGGTGCGGTGACAGGTCCC	1800
540	R V R H T I T I P N R G G A G A V T G P	559
1801	GACGAATGGGCGGACTTTGGCTTTGACCTGCGTACTGCAAGTCCCCGTAAGCAGCCCCATC	1860
560	D E W A D F G F D L P D C K S R K Q P I	579
1861	AAAGAGGAGTTCACAGAGACAGAGAGCCCTGAGGAACGTACCTTCTCTCTCTCTCTCT	1920
580	K E E F T E T E S H	599
1921	CTCTGTGAGAACTGCTCTTGAAGTGGGACCTGTTGGCTGTGCCACAGAAACCAGCAA	1980
1981	GGACCTTCTGCCGGATGCCATTCTGAAGGGAAGTCGCTCATGAACAACTCCCTCTTGG	2040

FIG.7

1	TGGTCCCGCTTCGACCAAGACTCCGGCTACCAGCTTGCGGGCCCCGCGGAGGAGGAGACC	60
61	CCGCTGGGGCTAGCTGGGCGACGCGGCCAAGCGGCGCGGGAAGGAGGCGGGAGGAGCG	120
121	GGGCCCCGAGACCCGACTCGGGCAGAGCCAGCTGGGGAGGCGGGGCGCGCTGGGAGCCA	180
181	GGGGCCCCGGGTGGCCGGCCCTCCTCCGCCACGGCTGAGTGCCCGCGCTGCCTTCCCGCCG	240
241	GTCCGCCAAGAAAGGCGCTAAGCCTGCGGCAGTCCCCCTCGCCGCCGCTCCCTGCTCCGC	300
301	ACCCTTATAACCCGCGCTCCCGCATCCAGGCGAGGAGGCAACGCTGCAGCCCAGCCCTCG	360
361	CCGACGCCGACGCCCGGCCGGAGCAGAATGAGCGGCAGCGTTGGGGAGATGGCCCAGAC	420
-8	M S G S V G E M A Q T	11
421	CTCTTCTTCTCCTCCTCCACCTTCGAGCACCTGTGGAGTTCTCTAGAGCCAGACAGCAC	480
12	S S S S S S T F E H L W S S L E P D S T	31
481	CTACTTTGACCTCCCCCAGCCCAGCCAAGGGACTAGCGAGGCATCAGGCAGCGAGGAGTC	540
32	Y F D L P Q P S Q G T S E A S G S E E S	51
541	CAACATGGATGTCTTCCACCTGCAAGGCATGGCCCAAGTTCAATTTGCTCAGCAGTGCCAT	600
52	N M D V F H L Q G M A Q F N L L S S A M	71
601	GGACCAGATGGGCAGCCGTGCGGCCCCGGCGAGCCCCCTACACCCCGGAGCAGCGCCAG	660
72	D Q M G S R A A P A S P Y T P E H A A S	91
661	CGCGCCCACCCACTCGCCCTACGCGCAGCCCAGCTCCACCTTCGACACCATGTCTCCGGC	720
92	A P T H S P Y A Q P S S T F D T M S P A	111
721	GCCTGTCATCCCTTCCAATACCGACTACCCCGGCCCCC	758
112	P V I P S N T D Y P G P	123

FIG. 8

```

- Name: sr-p70a-cos3      Len:   650  Check: 9661  Weight:  1.00
- Name: sr-p70b-cos3      Len:   650  Check: 3605  Weight:  1.00
- Name: sr-p70-ht29       Len:   650  Check:    85  Weight:  1.00
- Name: sr-p70c-att20     Len:   650  Check: 4072  Weight:  1.00
- Name: sr-p70a-att20     Len:   650  Check: 4204  Weight:  1.00

-//
-
-      1                               50
- sr-p70a-cos3      .....MAQ STTTSPDGGT TFEHLWSSLE PDSTYFDLPQ SSRGNNEVVG
- sr-p70b-cos3      .....MAQ STTTSPDGGT TFEHLWSSLE PDSTYFDLPQ SSRGNNEVVG
- sr-p70-ht29       .....MAQ STATSPDGGT TFEHLWSSLE PDSTYFDLPQ SSRGNNEVVG
- sr-p70c-att20     .....
- sr-p70a-att20     MSGSVGEMAQ ...TSSSSSS TFEHLWSSLE PDSTYFDLPQ PSQGTSEASG

-      51                               100
- sr-p70a-cos3      GTDSSMD.VF HLEGMTTSVM AQFNLLSSTM DQMSSRAASA SPYTPAHAAS
- sr-p70b-cos3      GTDSSMD.VF HLEGMTTSVM AQFNLLSSTM DQMSSRAASA SPYTPAHAAS
- sr-p70-ht29       GTDSSMD.VF HLEGMTTSVM AQFNLLSSTM DQMSSRAASA SPYTPAHAAS
- sr-p70c-att20     ...MCMGPVY ..ESLG...Q AQFNLLSSAM DQMGSRAAPA SPYTPAHAAS
- sr-p70a-att20     SEESNMMD.VF HLQGM..... AQFNLLSSAM DQMGSRAAPA SPYTPAHAAS

-      101                              150
- sr-p70a-cos3      VPTHSPYAQP SSTFDTMSPA PVIPSNTDYP GPHHFEVTFQ QSSTAKSATW
- sr-p70b-cos3      VPTHSPYAQP SSTFDTMSPA PVIPSNTDYP GPHHFEVTFQ QSSTAKSATW
- sr-p70-ht29       VPTHSPYAQP SSTFDTMSPA PVIPSNTDYP GPHHFEVTFQ QSSTAKSATW
- sr-p70c-att20     APTHSPYAQP SSTFDTMSPA PVIPSNTDYP GPHHFEVTFQ QSSTAKSATW
- sr-p70a-att20     APTHSPYAQP SSTFDTMSPA PVIPSNTDYP GP.....

-      151                              200
- sr-p70a-cos3      TYSPLLKKLY CQIAKTCPIQ IKVSAPPPPG TAIRAMPVYK KAEHVTDIVK
- sr-p70b-cos3      TYSPLLKKLY CQIAKTCPIQ IKVSAPPPPG TAIRAMPVYK KAEHVTDIVK
- sr-p70-ht29       TYSPLLKKLY CQIAKTCPIQ IKVSTPPPPG TAIRAMPVYK KAEHVTDVVK
- sr-p70c-att20     TYSPLLKKLY CQIAKTCPIQ IKVSTPPPPG TAIRAMPVYK KAEHVTDIVK
- sr-p70a-att20     .....

-      201                              250
- sr-p70a-cos3      RCPNHELGRD FNEGQSAPAS HLIRVEGNNL SQYVDDPVTG RQSVVVPYEP
- sr-p70b-cos3      RCPNHELGRD FNEGQSAPAS HLIRVEGNNL SQYVDDPVTG RQSVVVPYEP
- sr-p70-ht29       RCPNHELGRD FNEGQSAPAS HLIRVEGNNL SQYVDDPVTG RQSVVVPYEP
- sr-p70c-att20     RCPNHELGRD FNEGQSAPAS HLIRVEGNNL AQYVDDPVTG RQSVVVPYEP
- sr-p70a-att20     .....

-      251                              300
- sr-p70a-cos3      PQVGTEFTTI LYNFMCNSSC VGGMNR RPIL IITLETRDG QVLGRRSFEG
- sr-p70b-cos3      PQVGTEFTTI LYNFMCNSSC VGGMNR RPIL IITLETRDG QVLGRRSFEG
- sr-p70-ht29       PQVGTEFTTI LYNFMCNSSC VGGMNR RPIL IITLEM RDG QVLGRRSFEG
- sr-p70c-att20     PQVGTEFTTI LYNFMCNSSC VGGMNR RPIL VIITLETRDG QVLGRRSFEG
- sr-p70a-att20     .....

-      301                              350
- sr-p70a-cos3      RICACPGRDR KADEDHYREQ QALNESSAKN GAASKRAFKQ SPPAVPALGP
- sr-p70b-cos3      RICACPGRDR KADEDHYREQ QALNESSAKN GAASKRAFKQ SPPAVPALGP
- sr-p70-ht29       RICACPGRDR KADEDHYREQ QALNESSAKN GAASKRAFKQ SPPAVPALGA
- sr-p70c-att20     RICACPGRDR KADEDHYREQ QALNESTTKN GAASKRAFKQ SPPAIPALGT
- sr-p70a-att20     .....

```

FIG. 9

	351				400
- sr-p70a-cos3	GVKKRRHGDE	DTYYLQVRGR	ENFEILMKLK	ESLELMELVP	QPLVDSYR..
- sr-p70b-cos3	GVKKRRHGDE	DTYYLQVRGR	ENFEILMKLK	ESLELMELVP	QPLVDSYR..
- sr-p70-ht29	GVKKRRHGDE	DTYYLQVRGR	ENFEILMKLK	ESLELMELVP	QPLVDSYR..
- sr-p70c-att20	NVKKRRHGDE	DMFYMHVRGR	ENFEILMKVK	ESLELMELVP	QPLVDSYRQQ
- sr-p70a-att20
	401				450
- sr-p70a-cos3	QQQQLLQRPS	HLQPPSYGPV	LSPMNKVHGG	VNKLPSVNQL	VGQPPPHSSA
- sr-p70b-cos3	QQQQLLQRPS	HLQPPSYGPV	LSPMNKVHGG	VNKLPSVNQL	VGQPPPHSSA
- sr-p70-ht29	QQQQLLQRPS	HLQPPSYGPV	LSPMNKVHGG	MNKLPSVNQL	VGQPPPHSSA
- sr-p70c-att20	QQQQLLQRPS	HLQPPSYGPV	LSPMNKVHGG	VNKLPSVNQL	VGQPPPHSSA
- sr-p70a-att20
	451				500
- sr-p70a-cos3	ATPNLGPVGS	GMLNNHGHAV	PANSEMTSSH	GTQSMVSGSH	CTPPPPYHAD
- sr-p70b-cos3	ATPNLGPVGS	GMLNNHGHAV	PANSEMTSSH	GTQSMVSGSH	CTPPPPYHAD
- sr-p70-ht29	ATPNLGPVGP	GMLNNHGHAV	PANGEMSSSH	SAQSMVSGSH	CTPPPPYHAD
- sr-p70c-att20	AGPNLGPMSG	GMLNSHGHS	PANGEMNGGH	SSQTMVSGSH	CTPPPPYHAD
- sr-p70a-att20
	501				550
- sr-p70a-cos3	PSLVSFLTGL	GCPNCIEYFT	SQGLQSIYHL	QNLTIEDLGA	LKIQEYRMT
- sr-p70b-cos3	PSLVR..T.W	G.P.....
- sr-p70-ht29	PSLVSFLTGL	GCPNCIEYFT	SQGLQSIYHL	QNLTIEDLGA	LKIQEYRMT
- sr-p70c-att20	PSLVSFLTGL	GCPNCIECFT	SQGLQSIYHL	QNLTIEDLGA	LKVPDQYRMT
- sr-p70a-att20
	551				600
- sr-p70a-cos3	IWRGLQDLKQ	GHDYGAQAQ	LLR.SSNA	ISIGGSGELQ	RQRVMEAVHF
- sr-p70b-cos3
- sr-p70-ht29	IWRGLQDLKQ	GHDYS.TAQ	LLR.SSNA	ISIGGSGELQ	RQRVMEAVHF
- sr-p70c-att20	IWRGLQDLKQ	SHDCG...Q	LLRSSNA	ISIGGSGELQ	RQRVMEAVHF
- sr-p70a-att20
	601				650
- sr-p70a-cos3	RVRHTITIPN	RGGPGG..GP	DEWADFGFDL	PDCKARKQPI	KEEFTEAEIH
- sr-p70b-cos3
- sr-p70-ht29	RVRHTITIPN	RGGPGG..GP	DEWADFGFDL	PDCKARKQPI	KEEFTEAEIH
- sr-p70c-att20	RVRHTITIPN	RGGAGAVTGP	DEWADFGFDL	PDCKSRKQPI	KEEFTETESH
- sr-p70a-att20

FIG.9
suite

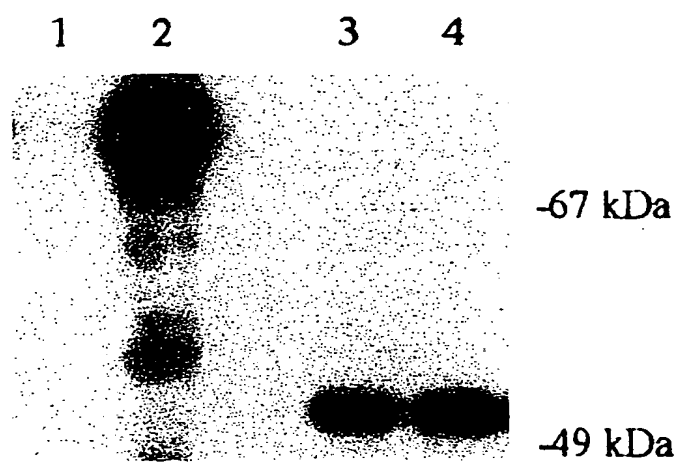


FIG.10a

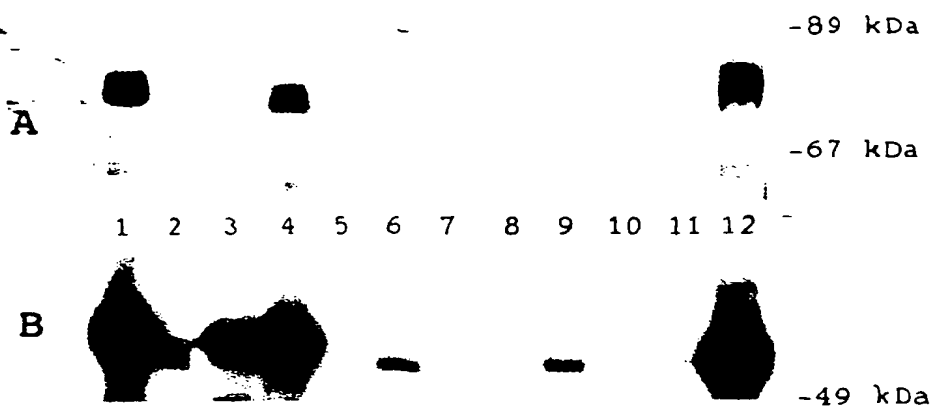


FIG.10 b

CONFIDENTIAL

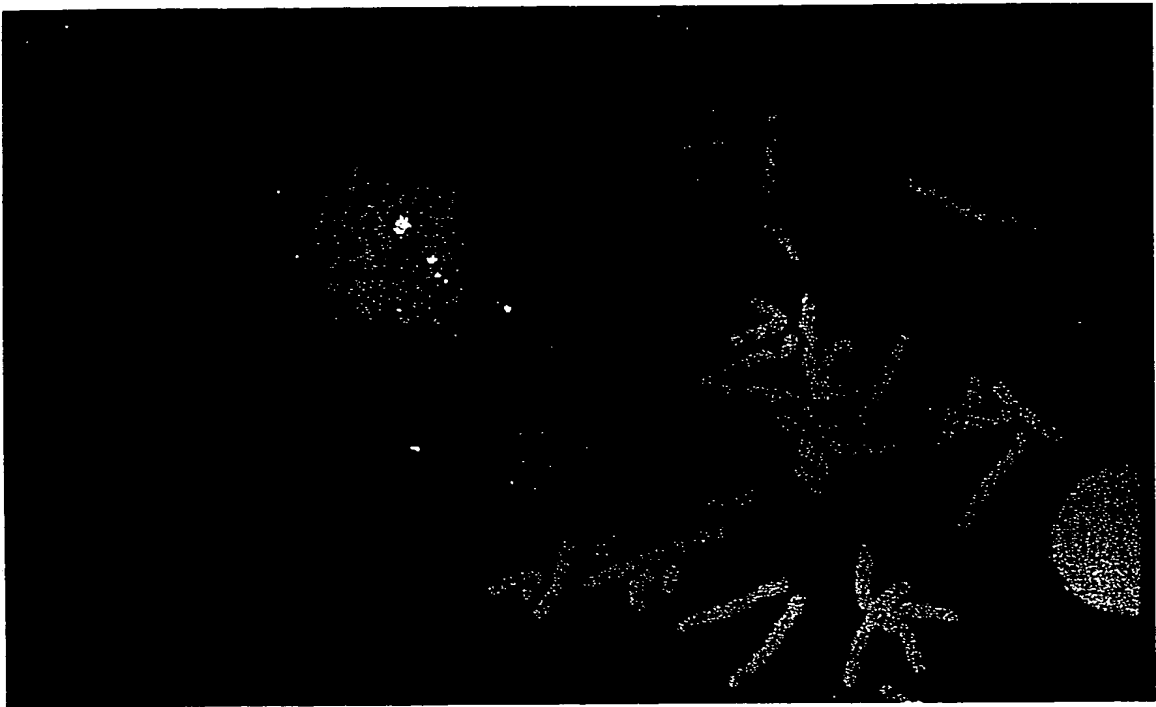
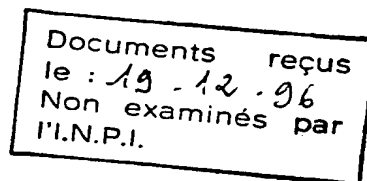


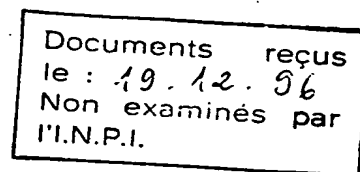
FIG.11

THIS PAGE BLANK (USPTO)



REVENDICATIONS

1. Polypeptide purifié, comprenant une séquence d'acides aminés choisie parmi :
 - a) la séquence SEQ ID n° 2 ;
 - b) la séquence SEQ ID n° 4 ;
 - c) la séquence SEQ ID n° 6 ;
 - d) la séquence SEQ ID n° 8 ;
 - e) la séquence SEQ ID n° 10 ;
 - f) toute séquence biologiquement active dérivée de SEQ ID n° 2, SEQ ID n° 4, SEQ ID n° 6, SEQ ID n° 8 ou SEQ ID n° 10.
2. Polypeptide selon la revendication 1, caractérisé en ce qu'il comprend la séquence d'acides aminés SEQ ID n° 6.
3. Polypeptide selon la revendication 1, caractérisé en ce qu'il comprend la séquence comprise entre :
 - le résidu 110 et le résidu 310 de SEQ ID n° 2, 4 ou 6 ;
 - le résidu 60 et le résidu 260 de SEQ ID n° 8 ;
 - le résidu 109 et le résidu 123 de SEQ ID n° 10.
4. Polypeptide selon la revendication 1, caractérisé en ce qu'il résulte d'un épissage alternatif de l'ARN messenger du gène correspondant.
5. Polypeptide selon l'une quelconque des revendications précédentes, caractérisé en ce qu'il s'agit d'un polypeptide recombinant produit sous la forme d'une protéine de fusion.
6. Séquence d'acides nucléiques isolée codant pour un polypeptide selon l'une quelconque des revendications précédentes.
7. Séquence d'acides nucléiques isolée selon la revendication 6, caractérisée en ce qu'elle est choisie parmi :
 - a) la séquence SEQ ID n° 1 ;
 - b) la séquence SEQ ID n° 3 ;
 - c) la séquence SEQ ID n° 5 ;



- d) la séquence SEQ ID n°7 ;
- e) la séquence SEQ ID n°9 ;
- f) les séquences d'acides nucléiques capables de s'hybrider à la séquence SEQ ID n° 1, SEQ ID n° 3, SEQ ID n° 5, SEQ ID n°7 ou SEQ ID n°9 ou à leurs séquences complémentaires, ou de s'hybrider à leurs séquences proximales.;
- g) les séquences dérivées des séquences a), b), c), d), e) ou f) du fait de la dégénérescence du code génétique.

- 8. Séquence nucléotidique selon la revendication 6, caractérisée en ce qu'il s'agit de la séquence SEQ ID n° 5 codant pour le polypeptide de séquences SEQ ID n° 6.
- 9. Vecteur de clonage et/ou d'expression contenant une séquence d'acides nucléiques selon l'une quelconque des revendications 6 à 8.
- 10. Vecteur selon la revendication 9, caractérisé en ce qu'il s'agit du plasmide pSE-1.
- 11. Cellule hôte transfectée par un vecteur selon la revendication 9 ou 10.
- 12. Cellule hôte transfectée selon la revendication 11, caractérisée en ce qu'il s'agit de *E. coli* MC 1061.
- 13. Sonde nucléotidique ou amorce nucléique caractérisée en ce qu'elle s'hybride spécifiquement avec l'une quelconque des séquences selon les revendications 6 à 8 ou leurs séquences complémentaires ou les ARN messagers correspondants ou les gènes correspondants.
- 14. Sonde selon la revendication 13, caractérisée en ce qu'elle comporte au moins 10 nucléotides.
- 15. Sonde selon la revendication 13, caractérisée en ce qu'elle comprend l'intégralité de la séquence du gène codant pour l'un des polypeptides de la revendication 1.

16. Sondes nucléotides ayant les oligonucléotides suivants ou leurs complémentaires :

SEQ ID n° 11 : GCG AGC TGC CCT CGG AG

SEQ ID n° 12 : GGT TCT GCA GGT GAC TCA G

SEQ ID n° 13 : GCC ATG CCT GTC TAC AAG

SEQ ID n° 14 : ACC AGC TGG TTG ACG GAG

SEQ ID n° 15 : GTC AAC CAG CTG GTG GGC CAG

SEQ ID n° 16 : GTG GAT CTC GGC CTC C

17. Utilisation d'une séquence selon l'une quelconque des revendications 6 à 8 pour la fabrication d'amorces oligonucléotidiques pour des réactions de séquençage ou d'amplification spécifique selon la technique de PCR ou toute variante de celle-ci.

18. Couple d'amorces nucléotidiques, caractérisé en ce qu'il comprend les amorces de séquences suivantes :

a) SEQ ID n° 11

amorce sens : GCG AGC TGC CCT CGG AG

SEQ ID n° 12

amorce antisens : GGT TCT GCA GGT GAC TCA G

b) SEQ ID n° 13

amorce sens : GCC ATG CCT GTC TAC AAG

SEQ ID n° 14

amorce antisens : ACC AGC TGG TTG ACG GAG

c) SEQ ID n° 15

amorce sens : GTC AAC CAG CTG GTG GGC CAG

SEQ ID n° 16

amorce antisens : GTG GAT CTC GGC CTC C

19. Utilisation d'une séquence selon l'une quelconque des revendications 6 à 8, pour la réalisation de sondes nucléotidiques de diagnostic ou de séquences antisens utilisables en thérapie génique.
20. Utilisation d'une sonde selon l'une quelconque des revendications 13 à 16, comme outil de diagnostic *in vitro* pour la détection, par des expériences d'hybridation, des séquences d'acides nucléiques codant pour un polypeptide selon l'une quelconque des revendications 1 à 4, dans des échantillons biologiques, ou pour la mise en évidence de synthèses aberrantes ou d'anomalies génétiques.
21. Méthode de diagnostic *in vitro* pour la détection de synthèses aberrantes ou d'anomalies génétiques au niveau des séquences d'acides nucléiques codant pour un polypeptide selon l'une quelconque des revendications 1 à 4, caractérisée en ce qu'elle comprend :
 - la mise en contact d'une sonde nucléotidique selon l'une quelconque des revendications 13 à 16 avec un échantillon biologique dans des conditions permettant la formation d'un complexe d'hybridation entre ladite sonde et la susdite séquence nucléotidique, éventuellement après une étape préalable d'amplification de la susdite séquence nucléotidique ;
 - la détection du complexe d'hybridation éventuellement formé ;
 - éventuellement le séquençage de la séquence nucléotidique formant le complexe d'hybridation avec la sonde de l'invention.
22. Utilisation d'une séquence d'acides nucléiques selon l'une quelconque des revendications 6 à 8, pour la production d'un polypeptide recombinant selon l'une quelconque des revendications 1 à 5.
23. Méthode de production d'un polypeptide selon l'une quelconque des revendications 1 à 5 recombinant, caractérisée en ce que l'on cultive des cellules transfectées selon la revendication 11 ou 12 dans des conditions permettant l'expression d'un polypeptide recombinant de séquence SEQ ID n° 2, SEQ ID n° 4, SEQ ID n° 6, SEQ ID n° 8 ou SEQ ID n° 10 ou tout fragment ou dérivé biologiquement actif, et que l'on récupère ledit polypeptide recombinant.

24. Anticorps mono ou polyclonaux ou leurs fragments, anticorps chimériques ou immunoconjugués, caractérisés en ce qu'ils sont capables de reconnaître spécifiquement un polypeptide selon l'une quelconque des revendications 1 à 4.
25. Utilisation des anticorps selon la revendication précédente, pour la purification ou la détection d'un polypeptide selon l'une quelconque des revendications 1 à 4 dans un échantillon biologique.
26. Procédé de diagnostic *in vitro* de pathologies corrélées à une expression ou une accumulation anormale de protéines SR-p70, notamment les phénomènes de cancérisation, à partir d'un prélèvement biologique, caractérisé en ce que l'on met en contact au moins un anticorps selon la revendication 24 avec ledit prélèvement biologique, dans des conditions permettant la formation éventuelle de complexes immunologiques spécifiques entre une protéine SR-p70 et le ou lesdits anticorps et en ce que l'on détecte les complexes immunologiques spécifiques éventuellement formés.
27. Kit pour le diagnostic *in vitro* d'une expression ou une accumulation anormale de protéines SR-p70 dans un prélèvement biologique et/ou pour la mesure du taux d'expression de celles-ci dans ledit prélèvement comprenant :
 - au moins un anticorps selon la revendication 24, éventuellement fixé sur un support,
 - des moyens de révélation de la formation de complexes antigènes/anticorps spécifiques entre une protéine SR-p70 et ledit anticorps et/ou des moyens de quantification de ces complexes.
28. Méthode pour le diagnostic précoce de la formation des tumeurs caractérisée en ce que l'on met en évidence dans un échantillon de sérum prélevé chez un individu des auto-anticorps dirigés contre une protéine SR-p70 selon les étapes consistant à mettre en contact un échantillon de sérum prélevé chez un individu avec un polypeptide de l'invention, éventuellement fixé sur un support, dans des conditions permettant la formation de complexes immunologiques spécifiques entre

ledit polypeptide et les auto-anticorps éventuellement présents dans l'échantillon de sérum, et en ce que l'on détecte les complexes immunologiques spécifiques éventuellement formés.

29. Composition pharmaceutique comprenant, à titre de principe actif, un polypeptide selon l'une quelconque des revendications 1 à 4.
30. Composition pharmaceutique selon la revendication précédente, caractérisée en ce qu'elle comprend un polypeptide selon la revendication 2.
31. Composition pharmaceutique contenant un inhibiteur ou un activateur de l'activité du SR-p70.